REVIEW ARTICLE

A Review of Artificial Intelligence–Driven Predictive Strategies for the Formulation Development of Oral Solid Dosage Forms

Monisha Alla[1*], Rajesh Vooturi[2], Veena Vithalapuram[3], Jagadeesh Nadigoti[4]

[1]Formulation Scientist, Department of Formulation Research & Development, Aurigene Pharmaceutical Services Limited (A wholly owned subsidiary of Dr. Reddy's Laboratories), Miyapur, Hyderabad - 500049, Telangana, India.

[2]Head of the Department, Department of Formulation Research & Development, Aurigene Pharmaceutical Services Limited (A wholly owned subsidiary of Dr. Reddy's Laboratories), Miyapur, Hyderabad - 500049, Telangana, India.

[3]Team Leader, Department of Formulation Research & Development, Aurigene Pharmaceutical Services Limited (A wholly owned subsidiary of Dr. Reddy's Laboratories), Miyapur, Hyderabad - 500049, Telangana, India.

[4]Specialist, Department of Formulation Research & Development, Aurigene Pharmaceutical Services Limited (A wholly owned subsidiary of Dr. Reddy's Laboratories), Miyapur, Hyderabad - 500049, Telangana, India.

Corresponding Author Email ID: monishaa@aurigeneservices.com

Received: 10 February 2026 **Accepted:** 30 March 2026 **Published:** 15 April 2026

Abstract

Artificial intelligence is continuously developing in the area of formulation research & development (R&D) by increasing predictive and data-driven approaches in designing and development of oral solid dosage (OSD) forms. Earlier, formulation development was largely dependent on multiple experimentations, which were time-consuming and required a substantial number of resources or scientists, which ultimately ended up in high development costs. AI-driven predictive modeling through machine learning, deep learning, and quantitative structure - property relationship (QSPR) techniques helps in rapid prediction of the critical formulation attributes (CFA's) like solubility, dissolution, stability, and bioavailability. By inserting inputs like the experimental data, API / excipient properties, and all the process parameters, these models predict the desired formulation optimization and accelerate prototype selection.

This review article summarizes the evolving role of AI in OSD formulation R&D, highlighting its application in drug excipient compatibility prediction, process optimization, scale-up for pilot and commercial scale batches and accelerated stability prediction. Also, a few points related to the challenges related to data quality, model intelligibility and regulatory acceptance are discussed. AI-driven predictive approaches hold significant promise to transform formulation R&D by enabling faster, smarter, and more efficient development of oral solid dosage forms.

Keywords: Artificial Intelligence, Machine Learning, Deep Learning, Predictive Modeling, Oral Solid Dosage Forms, Quality by Design, Critical Quality Attributes, Critical Material Attributes, Process Analytical Technology, Design of Experiments.

Journal of Applied Pharmaceutical Sciences and Research, (2026); DOI: 10.31069/japsr.v9i1.02

Introduction

Oral solid dosage forms (OSDs) are the most widely used by people of different age groups and the most accepted pharmaceutical dosage forms because of their several advantages. Almost 70 to 80% of the marketed drug products are Oral solid dosage forms, which include tablets, capsules, powders, and granules. Oral dosage forms are the pharmaceutical formulations that are designed and formulated for oral administration only, that are intended to deliver a precise dose of an active pharmaceutical ingredient (API) through the gastrointestinal (GI) tract, and they are classified as follows

- Immediate release (IR) dosage forms, which are designed for rapid drug release and absorption.
- Modified release (MR) forms like sustained release, controlled release, delayed release, etc., which are developed to optimize pharmacokinetic profiles.
- Specialized dosage forms like orally disintegrating tablets (ODTs), chewable tablets, multiparticulate systems, etc., which are aimed at improving patient acceptability and bioavailability.[1]

Oral solid dosage forms have a significant importance in the pharmaceutical industry as they are non-invasive, simple and acceptable across all age groups when compared to other dosage forms. Also, OSDs are easier and less expensive to manufacture, have a greater physical and chemical stability compared to the liquid or semisolid dosage forms, and are relatively easy to store and transport. Despite all these advantages traditional OSD's development was largely empirical and facing many challenges related to poor solubility, stability, permeability, manufacturing variability and patient compliance issues for a new drug molecule. The

emergence of modern predictive modeling, material science, and QbD approach has significantly mitigated many of these limitations, providing the way for more robust and efficient OSD design formulations. The near future research focuses on Novel excipients, other technologies like Hot-melt extrusion and 3D printing, In- silico modeling and AI-driven formulation design for predictive modeling and on personalized oral dosage forms tailored to individual therapeutic needs [2].

The term "Artificial Intelligence" is the development of computer systems and machines that can perform tasks normally requiring human intelligence, such as learning, reasoning, decision-making and perception, often using artificial neural networks. This Artificial intelligence is a subfield of computer science where it was developed in a way that addresses the problems in engineering, business, and Healthcare divisions. In the recent era, AI is gaining excellent value very rapidly across different sectors out of which the pharmaceutical industry is one of the leading sectors. In the past, formulation scientists have used traditional techniques for drug product development, which were time-consuming. The need for AI-driven predictive modeling and data-driven approaches in Orals solid dosage form development has gained a lot of importance because of its:[3]

- Reduce trial-and-error / empirical work, like predictive modeling has the ability to screen trials faster with lower material cost and fewer failed batches.
- Better predictability of critical quality attributes (CQA's) like the dissolution profiles, stability under various conditions, bioavailability, hardness, friability based on formula composition and processing it is easier to confirm quality, reduce batch variability and can be able to satisfy the regulatory expectations.
- Handle complexity & multivariate dependencies, for example, many variables like API solubility, particle size, excipient type, amount of excipient, compression force, moisture, etc., can affect the tablet performance. The interactions between these factors are mostly non-linear. Data-driven methods like machine learning and multivariate statistical models can provide these interactions better than simple OFAT studies.
- Cost & time effective because of its fewer experimentation, faster cycles, less material used altogether contribute to lowering the cost and in turn speed up the Research and Development.
- Quality by design (QbD) as most of the regulatory agencies expect a deep understanding of formulation, its process and product stability. AI - Predictive models and data-driven designs help in defining/setting the acceptable ranges, designing control strategies, and risk assessments
- Optimization and scale-up as formulations go from R&D scale to pilot/commercial scale. Based on the large-scale equipment to be used, many process parameters are bound to change. Predictive tools can help detect or identify how these changes can affect the key properties; in turn, the scale-up will have a successful outcome.
- Driving novel formulation complexities, such as most of the new APIs have poor solubility, polymorphism, and instability for such complex molecules. Data-driven tools will help to predict solubility enhancement and co-crystal formation for effective development.
- Personalization/Flexibility, some modern trends include flexible or personalized dosing like minitablets, multiparticulate systems, 3D printing, etc. To make such customizable formulations feasible, faster predictive approaches are needed to ensure the variant of the formulation meets quality, performance and/ safety criteria [3,4].

Artificial intelligence has rapidly grown as a game-changing technology across the different areas of pharmaceutical sciences, like in drug discovery, formulation development, manufacturing & quality control and in Clinical trials. With the arrival of AI technologies such as machine learning (ML) and deep learning (DL), the pharmaceutical sector has set foot in an era of accelerated innovation and precision. These technologies were used in the different areas of the pharmaceutical industries where they can process vast and complex datasets from preclinical research, formulation studies, and clinical trials to make predictive decisions that were previously not obtained through conventional methods.

In drug discovery and design, AI plays a key role in identifying drug targets, optimizing lead compounds, and predicting pharmacokinetic prediction and toxicity profiles. Various machine learning algorithms help in analyzing the large chemical and biological database for screening the potential molecules within a brief period of time, which reduces both time and cost in the early pharmaceutical development stage. AI-driven molecular modeling tools can predict the molecule/API binding affinities and their interactions, which are continuously improving the success rate of drug candidates stepping into the preclinical stage [5].

In the formulation development phase, especially for oral solid dosage forms, AI is providing invaluable support for predictive modeling and optimization. It provides scientist to predict solubility, stability, dissolution rates, and bioavailability based on molecular structures and formulation variables. By giving all the data from drug excipient compatibility (DEC) studies, API & excipient properties, all the process parameters, and *in-vitro* dissolution results, these AI-based models can suggest optimized formulation with minimum experiments, which is helpful not only in time reduction but also in completely aligning with the principles of Quality by Design (QbD) [6].

In the area of manufacturing and quality control, AI is transforming through various automation, process analytical technologies (PAT), and digital twins. All these tools allow

real-time monitoring of production parameters, predictive maintenance of equipment and deviation detections before they impact the product quality. The results obtained through these models are more efficient, consistent, and complete manufacturing processes, which also support the regulatory expectations for continuous process verification.

Also, in clinical research, AI helps in patient/ subject selection, Clinical trial design and data analysis. Predictive analytics helps in the identification of a suitable patient population, reducing trial failures and improving outcome predictability. AI algorithms can also analyze real-world evidence (RWE) from electronic health records and wearable devices, enhancing the understanding of treatment responses and safety profiles.[7]

Despite its enormous advantages and capacities, the integration of AI in Pharmaceutical sciences faces challenges such as data quality, transparency of algorithms, and regulatory acceptance. Also, some AI models are overly complex, like deep learning neural networks, which make predictions without clarity showing how they arrived at those results, where the internal reasoning process is not transparent or understandable. However, as regulatory agencies such as the FDA and EMA recognize that the value of AI – based tools is the greatest path towards the digital transformation in pharma is becoming very much clear.

This review completely focuses on the application of AI and predictive modeling approaches in the development of oral solid dosage forms, highlighting their role in enhancing formulation design, optimization, manufacturability and performance prediction. The main objective is to highlight how these AI-driven tools can accelerate decision-making, improve product understanding, reduce experimental workload, and support quality by design (QbD), while also meeting regulatory expectations. This review aims to present how the formulation scientists are effectively integrating AI into modern OSD development [8, 9].

Fundamentals of Artificial Intelligence and Predictive Modeling in Pharmaceutical R&D

Artificial intelligence refers to the computational-designed systems to provide or perform the tasks that typically require human intelligence, such as learning, reasoning, pattern recognition, and decision making. Within AI, machine learning (ML) represents a subset of methods that provide a systematic way of learning from the data and improving their predictions without explicit programming. Deep learning (DL) is a further specialized branch of machine learning that uses multi-layered neural networks to identify the complex, non-linear patterns, making it especially powerful for high-dimensional formulation and materials data.

There are diverse types of learning in Pharmaceutical predictive modeling sciences, like

Supervised learning

These are the models that learn from labeled datasets to predict CQAs such as dissolution rate, stability, disintegration time, or bioavailability. A few of the algorithms like random forest (RF), support vector machines (SVM), k-nearest neighbors (k-NN), and artificial neural networks (ANNs), are widely used for the formulation optimization.

Unsupervised Learning

Used to uncover the hidden patterns in material attributes, classification of excipient functionalities, and cluster formulation datasets, principal components Analysis (PCA), K-Means clustering, and Hierarchical clustering are the most commonly used tools.

Reinforcements Learning (RL)

Emerging area where models learn optimal formulation or process-control strategies through trial – and – reward mechanisms, aligning with continuous manufacturing and process optimization.

In oral solid dosage form development, many artificial intelligence and predictive models are applied for solubility enhancement, excipient screening, tabletability/ compressibility of blend into tablet, prediction, and process parameter optimization in each stage of formulation development. A few examples are Random Forest (RF), SVM, ANN/DL Models, PCA, Genetic Algorithms, Regression models (PLS, MLR), etc. All these AI and predictive models are used in *in-silico* experimentation, thereby reducing laboratory trials and accelerating formulation design compared to empirical, trial-and-error methods. However, data analytics and data engineering are considered the backbone of successful predictive modeling in pharmaceutical R&D. High-quality datasets should be curated from diverse sources, pre-formulation studies, process parameters, excipient libraries, PAT tools, and historical batch records. Data cleaning, feature selection, dimensionality reduction, and normalization are the foremost essential steps to ensure the model's reliability. Data engineering also facilitates integration of cheminformatics, PAT- based real-time data streams, and quality-by-design (QbD) frameworks, enabling robust, transparent, and regulatory-acceptable models. Ultimately, these practices improve the prediction accuracy, reduce variability in trials, and support decision making throughout the lifecycle of the formulation [9]

Predictive Modeling Approaches in OSD Formulation Development

Predictive modeling approaches in Material Property Prediction

Predictive modeling is one of the essential components of oral solid dosage formulation development, enabling early estimation of critical material attributes (CMAs)

AIdriven predictive modeling algorithms used for material property prediction in pharmaceutical development were summarized in Table 1.

In shot all these predictive modeling of CMAs strengthen the formulation-related decision-making by providing virtual experiments, reducing material consumption, and

Table 1: AIdriven predictive modeling algorithms used for material property prediction in pharmaceutical development

Category	*Examples*	*Relevance in material property predictions*
Statistical / QSPR/ Chemometric Models	Partial Least Squares (PLS), Principal Component Analysis (PCA), Multiple Linear Regression (MLR), Quantitative structure-property relationship (QSPR)	These models are commonly used to correlate molecular descriptors with physicochemical properties including solubility, permeability, LogP, pKa and Thermal transitions. These models allow rapid virtual screening of candidate molecules, helping to identify the most highly probable for development drug form and also, it predicts the wettability (highly / Poorly), or polymorphic risks before experimentation itself.
Deep Learning Models	DNN Graph Neural Networks Convolutional models	These models can learn intricate structure – property relationships, particularly for predicting dissolution relevant properties or polymorphic transformations based on spectral fingerprints (DSC, XRD, NIR) and molecular graphs.
Machine Learning Models	Random Forest (RF), Support vector Machines (SVM), Artificial Neural Networks (ANN), K-Nearest Neighbour (k-NN), Gradient boosting models (Eg: XGBoost)	All these Machine learning approaches provide higher predictive power for non-linear and multi-factorial material attributes that influence processability like predict flow properties, compressibility, bulk/tapped density, lubricant sensitivity, compatibility with excipients and moisture-related behavior. For example: ANN and SVM models have shown strong accuracy in estimating Heckle parameters, tensile strength, and Carr's index from particle size distribution and material morphology data, which allows the early detection of tableting or compressibility of blend in to tablets feasibility. All these Machine Learning model driven predictions support rational excipient selection and help forecast manufacturability risks, especially in direct compression formulation development systems.
Hybrid models	PBPK + ML, DOE + ML, Mechanistic + AI	These hybrid models which are in combination with machine learning and Artificial intelligence are mostly employed to link Critical material Attributes (CMAs) with Critical Quality attributes (CQAs) and in-vivo outcomes providing improved accuracy over individual models. These models / approaches also support digital workflow such as in-silico screening, formulation prototype selection, targeted experimentation, accelerating development cycles under QbD frameworks.

improving early risk assessment. Because of the continuous advancement of the high-quality datasets, PAT tools, and computational power, these modeling approaches are expected to grow toward more automated and mechanistically informed predictions, ultimately facilitating faster and more robust OSD development [10, 11].

Predictive modeling approaches in Formulation Optimization:

Machine learning models help to predict the critical quality attributes (CQAs) such as hardness, disintegration time, dissolution rate and the stability predictions based on the excipient ratios, API properties, and the process parameters like compression force, mixing time during granulation, granulation speed like impeller rpm, binder addition time and drying temperatures, etc.

- *Artificial Neural Networks (ANN)*

It captures the Non-linear relationship between the excipients and tablet properties, which in turn is useful for dissolution and drug release prediction.

For example, in the development of tablet dosage forms, the ANN models combined with design of experiments (DoE) data have been successfully used to define the binder and disintegrant concentrations in the formulation. This type of integrated modeling strategy significantly improves dissolution performance, in turn significantly reducing the number of experimental trials necessary for the optimization of formulation.

- *Random Forest (RF)*

This model is effective in identifying influential formulation variables and handling multivariate datasets with reduced overfitting.

For example, in the development of capsule dosage form, Random Forest models have been applied to identify critical excipient quantity or ratio that influences the powder flow and weight uniformity, thereby increasing overall process robustness and the product performance [9, 10].

- *Support Vector Machine (SVM)*

This model is suitable for regression and classification tasks, including formulation acceptability prediction.

Support Vector Machine (SVM) modeling can be successfully applied in the formulation development of immediaterelease tablets (IR), where binder concentration, disintegrant level, and compression force can be optimized to achieve the target tablet hardness and drugrelease (dissolution) profile. The experimental data generated through Design of Experiments (DOE) are used to prepare an SVM regression model, which enables the capture of the non-linear relationships between formulation variables and their critical quality attributes (CQAs) such as friability and dissolution (drug release kinetics).

- *Gaussian Process Regression (GPR)*

This model helps in the prediction of the design space and supports the exploration of formulation design space.

For example, during the wet-granulation process development of solid oral dosage forms like immediate release tablets or granules, Gaussian process regression, when integrated with Bayesian optimization, has facilitated precise control of the binder spray rate and impeller speed. This data-driven approach has improved the granule size distribution and the compressibility of the blend or granules while minimizing batch-to-batch variability from development to scale-up.

- *Gradient Boosting Algorithms*

This model offers high predictive accuracy for simple to complex multi-factor optimization problems.

The gradient boosting algorithms like XGBoost and LightGBM can be effectively applied in the formulation development of several types of tablet formulations, where the functional excipient ratios like binder, disintegrant, lubricant levels, and compression forces, such as precompression and main compression forces, are optimized to achieve the desired tablet hardness, dissolution profile and the friability parameters of the tablets. For this type of model, the experimental data generated through Design of Experiments (DoE) are used to train a gradient-boosting regression model, which gives a sequential decision tree that is constructed to minimize prediction error [12]. These modeling approaches record the complex non-linear relationship within the formulation and the process variables, where they identify the most critical factors in the formulation/process, and also predict the refined or optimized excipient combinations and the compression settings, like precompression and main compression setting, during the tableting process within the defined design space.

By combining the Design of Experiments (DoE) with artificial intelligence (AI) driven predictive models, with the structured experimental design and advanced predictive analytics, to significantly increase the formulation optimization. In the previous years, DOE has been performed to generate statistically robust and well-structured experimental data across the formulations and process design space. The data generated in the DOE is sequentially used to mentor AI-driven predictive models such as support vector machines, Gaussian process regression, and other gradient boosting algorithms as they capture the non-linear interactions over the potential capabilities of traditional quadratic models. This integrated technique increases the accuracy in the prediction of CQA's, efficient identification of optimal factor combinations, and significantly reduces the number of experiments, thereby supporting a Datadriven development strategy aligned with quality by design (QbD) principles [12, 13].

Predictive modeling approaches in process modeling and scale-up

Predictive modeling techniques are combined with pharmaceutical process development and scaleup, to authorize a data driven prediction of how every critical process parameter (CPP) like granulation, compression, coating etc. impacts the CQAs of the final dosage form, such as particle size distribution, moisture content, binder addition rate, impeller speed, tooling parameters, friability, hardness, weight gain and drug release behavior after coating. Advanced machine learning approaches combined with probabilistic modeling (probability theory), ensemble learning algorithms (combination of multiple techniques), and regressionbased methods (supervised machine learning techniques and statistical techniques) all these techniques are all being applied to anticipate process performance across the laboratory scale, like research and development work, pilot scale, and commercial scale manufacturing. By representing all the complex and multivariate correlations within process datasets, these models enable a quicker pharmaceutical process optimization, which in turn enhances the robustness of the final process that supports the more authentic technology transfer (TT), from R&D laboratory scale to pilot and commercial scale [13, 14].

Partial least squares regression (a statistical model that predicts the relation between a set of two variables) and principal component analysis (a machine learning technique used for transforming a large set of variables into smaller models are widely used multivariate statistical techniques in the process workflow design modeling in the pharmaceutical industry. These tools help to reveal the complex relationship between the critical formulation variables and process parameters by decreasing the dimensionality and by extracting the underlying and unobservable factors from multifeatured datasets. Despite all these primitive linear methods like the PLS and PCA models successfully predicting the majorly correlated process parameters data, and the data set is considerably used to observe, replicate, and optimize the pharmaceutical manufacturing processes.

For scale-up, predictive models help translate small-scale experimental findings to larger manufacturing scales by identifying scale-dependent parameters such as mixing time, shear rate, heat transfer and mass transfer coefficients. Mechanistic models (mathematical representations of a system that describe its behavior based on underlying physical, chemical, and biological principles) are widely linked with data-driven models, like hybrid models, to improve the uncertainty of estimation. By integrating these modeling approaches, they reduce the trial-and-error experimentation in the R&D scale that minimizes the risk of the process during the technology transfer (TT) that supports the quality by design and regulatory compliance.

Basically, during the development of granulation processes, the oral solid dosage form development, predictive modeling techniques are widely used to establish correlations between material attributes such as particle size distribution, moisture content, and binder concentration and process variables, including impeller speed, binder addition rate, and granulation time. All these inputs strictly correlate or are linked to critical granule parameters like

granule size distribution in the blend, granule density (bulk and tapped density), and flowability (granule flow property of the final blend). Few of the advanced AI driven predictive methods like ANNs, support vector machines (SVMs), and partial least squares (PLS) regression, were shown to have a strong ability in representing and generating a complex non-linear interaction that occur during several types of granulating methods such as granulation using high-shear/ rapid mixer granulation and the fluid-bed granulation in Fluid bed processor[15]. These modeling techniques provide an increasingly accurate prediction and in-depth granulation process understanding by any of the granulating methods, which in turn supports the QbD approaches and provides more robust process controls throughout process development.

During the technology transfer of finished products from lab scale to pilot or commercial scale, there will be a chance to face challenges during the unit operation of compression. In order to minimize or to avoid those challenges, various types of AI-driven predictive modeling techniques were used, such as SVM, ANN, Gaussian process regression (GPR) and parallel artificial neural networks (P-ANN). These AI driven predictive modeling techniques were used to focus on tableting parameters of the blend/granules like compaction forces like main compression force and precompression forces, dwell time (the duration that the tablet punch remains in full contact with the roller, indicating maximum force is applied to the product), lubrication level and tooling parameters to predict the tablet hardness, friability of compressed tablet, porosity after compression and distribution behavior of blend. All these data-driven models help in understanding all the scale-dependent parameters, such as press speed and die and hopper design, allowing an effective technology transfer from pilot to commercial tablet presses (different makes and models of compression machines) by minimizing the re-optimization of product [16].

In film coating, AI driven predictive modeling techniques are employed to determine the coating thickness like weight build up, coating uniformity on the tablet surface, %weight build up, and the consistent drugrelease behavior based on critical coating process parameters such as the % solid content of the coating dispersion, spray rate during the coating process, atomization pressure, inlet temperature, and pan speed. All these models notice the complex interactions between droplet formation during spraying the dispersion, drying dynamics to prevent tableting defects like twinning and sticking, polymer coalescence, and tablet bed movement to ensure uniform distribution of coating material on the tablet bed, authorize to more powerful control of coating quality. The combination of computational fluid dynamics (CFD) with AI-driven predictive modeling techniques and machine learning (ML) algorithms improves the capacity to predict coating efficiency, such as droplet spray pattern, heat and mass transfer, and drying kinetics[17, 18]. Such hybrid modeling approaches are especially valuable during scaleup from laboratoryscale coaters to fullscale production equipment, where airflow patterns, nozzletobed distances, and thermal conditions differ significantly. By using these predictive tools, formulation scientists can optimize process parameters, which in turn minimize experimental trials, to ensure coating uniformity and performance consistency across all scales, starting from development to commercial scales[9].

Application of AI in continuous manufacturing and Process Analytics Technology (PAT)

AI is playing a most noticeable role in the continuous manufacturing of solid oral dosage forms in the pharmaceutical industry. Advanced algorithm models can process large volumes of data generated from process analytical technology (PAT) tools such as near infrared spectroscopy (a rapid and non-destructive technique using infrared light to measure molecular vibrations), Raman spectroscopy (a non-destructive chemical analysis technique that identifies material and analyzes molecular structures based on the inelastic scattering of laser light), and inline particle-size analyzers. Machine learning models like K–nearest neighbour (k-NN), gradient boosting models, random forest (RF), SVM, and ANN use all the data to predict critical product attributes such as blend uniformity, moisture content, granule growth properties during wet granulation, and tablet quality during continuous operations in larger commercial scales, all these parameters are the critical pharmaceutical quality control process during the product development. The combination of artificial intelligence with both continuous direct compression and wet-granulation process reveals the prediction of process adjustments during the development and across the scales to key parameters such as feeder rates, blender speed, and compression forces. This capability improves the process robustness, which in turn minimizes the API/material wastage and enables faster scale-up with very few trials. Eventually, artificial intelligence-driven predictive modeling techniques have become the most important tools in the process controls of efficiency improvements, consistency in the process, and consistency across the continuous manufacturing platforms.[20]

Real-time monitoring and controls using Artificial Intelligence (AI)

Real-time monitoring and controls using Artificial intelligence are playing a very crucial role in the pharmaceutical industry, where real-time monitoring and controls promote the continuous manufacturing and supply in the pilot and commercial scales. AI-driven real-time monitoring controls provide a control that helps to ensure the product is manufactured with a desired quality. Using several types of AI-driven methods, such as Advanced machine learning techniques like Reinforcement learning (RL) and long short-term memory (LSTM) networks, used to predict the maintenance of equipment and other predictive techniques like Deep Neural Networks (DNN), SVM, PLS, and Hybrid mechanistic, along with machine learning techniques are

used to predict real-time process controls. Also, multivariate models are frequently used for extracting process knowledge from the data provided by the process analyzers, such as soft sensors and spectroscopic measurements. For example, during the large-scale manufacturing process, variations in the blend uniformity can be detected using the spectroscopic methods (NIR spectroscopy) and can provide an automatic signal for feeder adjustments, which in turn maintains material consistency and composition by preventing variability in the final finished product.

Process analytical technology (PAT) can be applied as a control during continuous manufacturing, which measures the quality attributes of the final finished product. All the artificial intelligence-enabled control strategies facilitate the continuous manufacturing, naturally providing a real-time release testing (RTRT) as they have the ability to evaluate and ensure product quality during the in-process and final product testing.

Multivariate statistical process control (MSPC) systems like Principal Component Analysis (PCA) and Partial Least Square (PLS) are used for the process monitoring during the large-scale manufacturing to determine the variability in the process over time and can detect the abnormal process events that may lead to Equipment malfunction during the manufacturing process, process deviations based on equipment related issues or process related and out of specification in the finished product [21].

AI-driven Supervisory Control and Data Acquisition (SCADA) technique is a combination of hardware and software frameworks that monitor, control and analyze the data from equipment in real time. It can measure the process parameters, incoming raw materials and also in-process material attributes, as well as final product quality attributes within the model that contributes Real time release and testing.

Integration of AI with Quality by Design (QbD)

Integration of artificial intelligence (AI) with quality by design (QbD) increases systematic pharmaceuticals development by transforming data-driven insights into predictive knowledge frameworks. As mentioned in the ICH guidelines, particularly in Q8(R2), Q9 and Q10, QbD highlights predefined objectives, process understanding and risk management. AI complements this framework by permitting advanced data mining, multivariate analysis, and real-time decision making.

Artificial intelligence-driven predictive models and QbD approaches combinedly focus on multivariate predictive modeling of formulation variables and process variables, which allows in-depth interpretation of complex interactions those influencing the final finished product quality. These systems facilitate early detection of high-risk parameters, which allows the proactive control strategies before the variability impacts the performance[22]. By increasing the use of adaptive learning mechanisms, Artificial Intelligence is continuously processing the models based on new processed data, which supports the ongoing process changes and improvements to be done for effective product development. Additionally, using the data generated through the AI-driven predictive modeling techniques during the development enhances the regulatory approval confidence, which in turn satisfies the upcoming scientific expectations and high-quality product development.

Design space development through predictive modeling

Design space in quality by design is a multidimensional, scientifically established combination of input variables like raw materials attributes and process variables like critical process parameters that establish a consistent and high-quality pharmaceutical product. In the past few years, design space development has shown a significant enhancement using the artificial intelligence-driven predictive models. By combining the developmental batch data and the process analytical technology (PAT) data or inputs, these AI-driven predictive models can generate a true and a scientific based design space rather than a normal one. This design space development approach is especially important and useful, particularly in the different unit operations during the manufacturing, such as the granulation stage, compression stage, and coating stages. Where this design stage development is more useful in the continuous manufacturing processes, where all the parameter interactions are more complex and highly interdependent.[23]

Increasing Process Understanding and Risk Assessment using AI tools

Process understanding and risk assessment analysis are the key steps during the formulation development of any oral solid dosage form. Initial risk assessment provides a path for the formulation development. In recent years, artificial intelligence has significantly enhanced the process of understanding by identifying the hidden key parameters that may not be noticeable during the day-to-day development. AI driven predictive modelings such as principal component analysis (PCA), clustering algorithms (these are group of unsupervised machine learning techniques that groups unlabeled data into clusters based on similarities in their features, such as density), and deep learning techniques helps to map a relationship between process variables and product performance during the early development stage which enable a risk free and efficient process technology transfer from small scale to commercial scales. All these AI-driven predictive modeling risk assessment tools were aligned with ICH Q9 principles. As discussed previously, a few of the supervised machine learning models can predict the Quality attributes like OOS of the finished product based on the early-stage process indicators, which helps in the prevention of adjustments during the manufacturing of the product. In the same way, Artificial intelligence enables real time monitoring system that helps to support adaptive controls during manufacturing, which in turn minimizes

batch failure and helps in delivering the product with high quality[18, 24].

Data Sources and Digital Infrastructure for AI in Formulation R&D:

Artificial intelligence-driven predictive modeling techniques used in the formulation R&D mostly depend on grouped data generated during the development, high-quality, and strong digital infrastructure. Extensive data sources from the R&D, including Preformulation studies, manufacturing process parameters, outputs from the process analytical techniques (PAT), stability data of the final product, and all the developmental batch records, were combined to machine learning models for the prediction of CQAs that helped in optimizing the final formulation. Effective data generation, standardization of the process, and the creation of consistent data are essential to ensure data integrity (DI), reproducibility and all the practices that are in line with regulatory compliance[11, 26].

A few of the digital platforms, like electronic lab notebook (ELN), and Laboratory Information Management system (LIMS) provide the feasibility to capture structured and a real time data, time- to-time traceability and real-time accessibility to all the readers and reviewers, which promotes model development and validation. Moreover, all the formulations, manufacturing process related data captured using these digital platforms like ELN & LIMS used in the different artificial intelligence driven predictive modeling techniques allows integrated *in-silco* experimentation (research performed using computer simulation, analytical techniques and modeling techniques), risk assessment using the available data, exploration of design space, significantly contributing to the development of formulation by reducing the API/ Material usage and reduces the number of experimentations.

Advances in AI – Assisted formulation development:

Advances in artificial intelligence-assisted formulation development include machine learning models, deep learning and predictive modeling across the different stages of pharmaceutical development. Few models across the stages were presented in the table below: [27, 28]

AIdriven predictive modeling algorithms used across oral solid dosage (OSD) formulation development and their roles in development were summarized in Table 2.

All the above-mentioned tools/models/algorithms are successful AI-assisted predictive tools and platforms that are successfully used in the various stages of pharmaceutical development, which help in faster and more effective development. Also, all these artificial intelligence-driven predictive models, which are used in the industry, have now been adopted by academic research centers, which has strengthened the collaboration between the academic research and pharmaceutical industry by combining the large experimental datasets from laboratory experiments to pharmaceutical companies, enables to formulate a quality and safety product for the patients. By using this artificial intelligence-driven predictive modeling across the stages of formulation development predicts the most critical quality attributes of the finished dosage form, such as tablet hardness, dissolution, and stability. Combining the data generated during the R&D development with these algorithms or models enables to formulate a quality product in line with all regulatory compliance.[29, 30].

Table -2: AIdriven predictive modeling algorithms used across oral solid dosage (OSD) formulation development and their roles

Stage: Pre-formulation & API Property Prediction		
Objective	*Prediction*	*Common AI/ML algorithms*
Physicochemical & Biopharmaceutical profiling	Solubility, permeability, LogP, Stability, pKa, Salt selection	QSAR models, Random Forest (RF), Support vector regression (SVR/ SVM), Partial Least squares (PLS), Artificial Neural Networks (ANN), Gaussian process regression (GPR), K-Nearest Neighbors (K-NN)
Early developability assessment	BCS class, Hygroscopicity, crystallinity behavior	Decision trees (DT), Bayesian models, clustering (k-means)
Stage: Excipient compatibility & selection		
Objective	*Prediction*	*Commonly used algorithms*
Excipient screening and interaction prediction	API-excipient compatibility, synergistic or antagonistic behavior	Random Forest, naive Bayes, Association Rule Mining, SVM, Clustering
Excipient ratio optimization	Solid-state stability & incompatibility	Genetic Algorithms (GA), ANN, DOE + ML hybrids.
Stage: Formulation design & optimization (e.g., Tablets, Capsules, granules etc)		
Objective	What is optimized	Commonly used algorithms
QbD based optimization	% Binder, % Disintegrant, Lubricant level, Granulation endpoint	Artificial Neural Networks Artificial Neural Networks (ANN), Response surface Methodology (RSM), Genetic Algorithms (GA), Support vector regression (SVR)

Multi-objective optimization	Dissolution, hardness, friability, DT, compressibility	NSGA-II, Particle swarm optimization (PSO), Bayesian Optimization
Stage: Process development (Granulation, Compression, Coating – PAT Integration)		
Process	AI role	Algorithms used
Wet / Dry granulation end-point prediction	Granule Size, LOD, Flow	Partial Least square (PLS), Artificial Neural Networks (ANN), Random Forest (RF)
Tableting (compression)	Predicting capping, hardness, tensile strength, sticking	SVM, ANN, GPR, process ANN (p-ANN)
Coating uniformity (PAT + ML)	Coating Thickness, surface defects	Computer Vision (CNN), Partial Least Square -DA, Principle Component Analysis, k-means clustering
Stage: Dissolution, release kinetics & IVIVC		
Objective	*Output*	*Algorithms*
Fast prediction of dissolution curves	% release vs time	ANN, Random Forest, SVR, PLS, Deep Learning (LSTM)
IVIVC Level A/B modeling	Plasma profile prediction	Non-linear ANN, Physiological DL models, Bayesian IVIVC models.
Stage: Stability & shelf-life modeling		
Objective	*Predictions*	*Algorithms*
Stability under ICH conditions	Degradation kinetics, Shelf life	Time-series ML models (LSTM, ARIMA-ML hybrids), RF, Gradient Boosting Machine (GBM), XGBoost
Forced degradation trend analysis	Impurity growth	Principle Component Analysis, Partial Least Square, Bayesian forecasting
Stage: Manufacturing scale-up & continuous manufacturing		
Objective	*Algorithms*	
Predictive maintenance of equipment	Reinforcement Learning (RL), Long Short-Term Memory (LSTM)	
Real-time process control (PAT-ML-QbD)	Deep Neural Networks (DNN), Support Vector Machine (SVM), PLS, Hybrid Mechanistic + Machine Learning Models	
Stage: Quality control, image analysis & defect detection		
Area	*Algorithms*	
Tablet defects, Coating defects, Particle size	Convolutional Neural Network (CNN is a Deep Learning model), Machine Vision models, k-means, Principle Component Analysis	

Challenges and Limitations

Despite the rapid evolution of AI–driven predictive modeling and its growing adoption in pharmaceutical dosage form development, several challenges and limitations continue to restrict its broader implementation and reliability.

Limited access to highquality data

Early development stages often involve minimal API availability and strict confidentiality constraints. As a result, datasets tend to be inconsistent, incomplete, and unreliable, reducing model robustness and limiting predictive accuracy.

Use of limited or overly generalized models

Many existing models fail to capture complex, domain-specific formulation and process nuances. Such oversimplification can lead to suboptimal or incorrect predictions, limiting trust and practical applicability.

Challenges in data integration and standardization

Pharmaceutical data originate from diverse instruments, formats, and experimental conditions. Variability across these sources complicates dataset harmonization, making it difficult to perform reliable, large-scale analytics.

Limited model interpretability and transparency

Advanced algorithms—particularly deep learning approaches—often function as "black boxes." Difficulty in understanding how predictions are generated can hinder scientific acceptance and slow regulatory adoption.

Regulatory and compliance barriers

Evolving regulatory expectations for AI/ML models create challenges in validating, documenting, and deploying such systems. Ensuring traceability, auditability, and reproducibility remains a key barrier.

High implementation costs and infrastructure requirements

Establishing AI capabilities demands significant investment in computational resources, secure data platforms, and specialized software, which may be prohibitive for smaller organizations.

Knowledge gaps and skill shortages

Developing, validating, and maintaining predictive models requires multidisciplinary expertise spanning formulation science, data science, statistics, and machine learning, a combination not always readily available.

Data privacy and intellectual property risks

Managing sensitive or proprietary datasets introduces concerns regarding confidentiality, ownership, and secure usage. Ensuring compliance with dataprotection frameworks is critical to mitigate these risks.[31,32]

Future Prospects

The future of artificial intelligence–driven predictive modeling techniques in pharmaceutical formulation development is expected to advance significantly with the rise of generative AI and multi-objective optimization, enabling smarter, faster, and more efficient finished-product design. Models such as GastroPlus, DD-Solver AI extensions, MSTAR, and Simcyp PBPK serve as domain-specific AI-enabled predictive tools that support accurate prediction of powder behavior during the compression, granulation performance, like granule growth, tableting characteristics, like hardness, friability, and dissolution profiles. Other AI-powered platforms, such as MODDE Pro, Design Expert with ML modules, MATLAB AI toolbox, KNIME Pharma extensions, and tableting predictive tools, support formulation final formula selection screening, DoE optimization by creating the design space, and excipient selection by predicting the compatibility between the API and Excipients[33, 34].

Also, few other advanced technologies such as predictive digital twin technologies such as Siemens Simcenter for tablet pressing and powder flow, Ansys Rocky DEM, AspenTech Process Modeling, Ansys Fluent, and COMSOL multiphysics provide real-time simulation of powder dynamics, granulation, drying, and compression which are also helps in continuous manufacturing for the effective manufacturing of the quality and regulatory compliance finished product. Independent formulation laboratories or academic research centers, equipped with AI-driven robotic platforms such as Chemspeed, Freeslate, and AI-enabled ELN/LIMS tools like LabVantage, enable automated experimentation and high-throughput formulation optimization, greatly accelerating development cycles [34, 35].

Artificial intelligence also opens new pathways for personalized oral solid dosage formulations through technologies such as AI-guided 3D printing, machine-learning-based dose tailoring, and patient-specific release modeling, allowing formulations to be tailored to individual therapeutic and physiological needs. In the same way, regulatory agencies are transitioning toward digital-quality and model-informed frameworks, supporting safer, more transparent, and broader adoption of AI-driven tools across pharmaceutical development.[36]

Conclusion

Artificial intelligence–driven predictive approaches are rapidly transforming oral solid dosage (OSD) formulation development by enabling dataguided formulation design, accelerated process optimization, and improved product quality. Albased predictive models—including machine learning and deep learning algorithms—when integrated with data generated through QbD studies and DoE, provide powerful tools for predicting optimal excipient ratios, excipient–API interactions, critical process parameters, and overall product performance. Furthermore, digital infrastructures such as LIMS and ELN strengthen these models by supplying structured, highquality data essential for accurate predictions. Despite these advancements, challenges remain in areas such as data quality, model interpretability, regulatory acceptance, and integration into existing development workflows. Future progress will depend on the availability of highquality, standardized data, as well as skilled data scientists within the pharmaceutical industry, and the development of transparent, regulatoryaligned AI models. Overall, AIdriven predictive modeling holds significant promise for accelerating formulation development, reducing experimental burden, and enabling more efficient and robust manufacturing of oral solid dosage forms.

Acknowledgment

The authors sincerely acknowledge the support and encouragement provided by the Formulation Research & Development Department at Aurigene Pharmaceutical Services (a subsidiary of Dr. Reddy's Laboratories) throughout the development of this review article. The team's constructive discussions, technical inputs, and collaborative spirit significantly contributed to enhancing the quality of this work. The authors are also grateful to the organization for providing access to scientific resources, digital tools, and a conducive research environment, all of which were essential for completing this review.

Funding

The authors declare that this research did not receive any grant or financial support from funding agencies in the public, commercial, or notforprofit sectors.

Declaration of Competing Interest

The authors affirm that there are no competing commercial interests or personal relationships that could have appeared to influence the work reported in this study.

Data Availability

The authors confirm that no datasets were generated or analyzed in conducting this review.

Declaration of Generative AI

Alassisted tools were utilized solely for language refinement and structural organization in the development of this manuscript. The authors independently verified, edited, and approved all sections of the final content.

References

1. Iqubal MK, Singh PK, Shuaib M, Iqubal A, Singh M. Recent advances in direct compression technique for pharmaceutical tablet formulation. *Int J Pharm Res Dev*. 2014;6:49–57.
2. He S, Liu Z, Xu D. Advance in oral delivery systems for therapeutic protein. *J Drug Target*. 2019;27:283–291. doi:10.1080/1061186X.2018.1486406.
3. Ali A, et al. Future Journal of Pharmaceutical Sciences. 2024. doi:10.1186/s43094-024-00625-1.
4. Vyas M, Thakur S, Riyaz B, et al. Artificial intelligence: the beginning of a new era in pharmacy profession. *Asian J Pharm*. 2018;12:72–76. doi:10.22377/ajp.v12i02.2317.
5. Rudrurkar MN, et al. The impact of artificial intelligence on drug discovery: A comprehensive review. *Asian J Pharm*. 2024;12(2):171–178. doi:10.22270/ajprd.v12i2.1230.
6. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ*. 2016;47:20–33. doi:10.1016/j.jhealeco.2016.01.012.
7. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019;18(6):463–477.
8. Bergström CAS, et al. Early pharmaceutical profiling to predict oral drug absorption. *Adv Drug Deliv Rev*. 2016;101:6–20.
9. Polak S, et al. PBPK modeling in drug development and regulatory science. *CPT Pharmacometrics Syst Pharmacol*. 2020;9(3):174–182.
10. Bender A, Cortés-Ciriano I. Exploring AI in pharma and drug discovery. *Drug Discov Today*. 2021;26(5):1202–1215.
11. Yu LX, et al. QbD and product development. *AAPS J*. 2014;16(4):771–783.
12. Toropov A. QSPR/QSAR in pharmaceutical property prediction. *Eur J Pharm Sci*. 2017.
13. O'Shea JP, et al. Chemometric modeling for physicochemical property prediction. *J Pharm Biomed Anal*. 2020.
14. Mosharraf M, Nyström C. Predicting flow and compressibility using machine learning. *Int J Pharm*. 1995.
15. Üstün B, et al. ANN and SVM for compression and tableting behavior. *Eur J Pharm Biopharm*. 2007.
16. Paixão P, et al. ML-based prediction of hygroscopicity and powder behavior. *Powder Technol*. 2019.
17. Yang K, et al. Deep learning for molecular and material property prediction. *J Chem Inf Model*. 2019.
18. Dokoumetzidis A, Macheras P. Hybrid and PBPK-AI modeling in formulation prediction. *Pharm Res*. 2019.
19. Han R, Yang Y, Dietrich J, Luebbert C, Jha SK. Machine learning approaches in pharmaceutical development and manufacturing: current status and future perspectives. *Int J Pharm*. 2021;599:120443.
20. Reklaitis GV, Khinast JG, Muzzio FJ. Pharmaceutical process design and optimization using experimental design and machine learning. *Annu Rev Chem Biomol Eng*. 2010;1:239–261.
21. García-Muñoz S, MacGregor JF. Machine learning in pharmaceutical process monitoring and optimization. *AIChE J*. 2014;60(6):2015–2024.
22. Boukouvala F, Floudas CA. Machine learning models for pharmaceutical formulation design and optimization. *Comput Chem Eng*. 2017;106:617–631.
23. Patel S, Patel NM. Statistical design of experiments for pharmaceutical formulation development. *J Pharm Innov*. 2008;3(2):113–122.
24. International Council for Harmonisation (ICH). ICH Q8(R2): Pharmaceutical development. Geneva: ICH; 2009.
25. Wu H, Khan MA. Quality-by-design (QbD): an integrated approach to pharmaceutical development of solid oral dosage forms. *J Pharm Sci*. 2012;101(3):1096–1106.
26. Fonteyne M, Vercruysse J, De Leersnyder F, Van Snick B, Vervaet C, Remon JP, et al. Process analytical technology for continuous manufacturing of solid-dosage forms. *TrAC Trends Anal Chem*. 2015;67:159–166.
27. Peeters E, Fonteyne M, Vervaet C, Remon JP. Scale-up of pharmaceutical granulation processes: predictive modeling and PAT tools. *Eur J Pharm Biopharm*. 2015;95:79–92.
28. Ierapetritou MG, Muzzio FJ. Process modeling and simulation of pharmaceutical continuous manufacturing of solid oral dosage forms. *Annu Rev Chem Biomol Eng*. 2016;7:123–146.
29. Pauli V, Roggo Y, Pellegatti L, Marini F. The potential of multivariate data analysis in PAT for solid oral dosage manufacturing. *J Pharm Biomed Anal*. 2019;164:166–178.
30. Van Snick B, Holman J, Vanhoorne V, Kumar A, De Beer T, Remon JP, et al. Development of a continuous direct compression platform using design of experiments and predictive modeling. *Int J Pharm*. 2017;529(1–2):394–405.
31. Roggo Y, Chalus P, Maurer L, Lema-Martinez C, Edwards D, Jent N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J Pharm Biomed Anal*. 2007;44(3):683–700.
32. Lee SL, O'Connor TF, Yang X, Cruz CN, Chatterjee S, et al. Modernizing pharmaceutical manufacturing: from batch to continuous production. *J Pharm Innov*. 2015;10(3):191–199.
33. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. *Drug Discov Today*. 2021;26(1):80–93.
34. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019;18(6):463–477.
35. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. *Drug Discov Today*. 2018;23(6):1241–1250.
36. Ierapetritou M, Muzzio F, Reklaitis G, et al. Process modeling and artificial intelligence applications in pharmaceutical manufacturing and development. *Comput Chem Eng*. 2020;134:106706.

How to cite this article: Alla M, Vooturi R, Vithalapuram V, Nadigoti J. A Review of Artificial Intelligence–Driven Predictive Strategies for the Formulation Development of Oral Solid Dosage Forms. Journal of Applied Pharmaceutical Sciences and Research. 2026; 9(1)10-20 Doi : 10.31069/japsr.v9i1.02