Emerging AI Tools For Developing Solid Dosage Forms: Current Scenario and Problems

Ms. Rohini Nale[1], Mr. Chandrakant Wadile[2], Dr. Pankaj Mandpe[2], Ms. Divyanka Bodas[2], Ms. Nidhi Kate[2]

[1]Department of Pharmaceutical Quality Assurance, H. R. Patel Institute of Pharmaceutical Education and Research, Shirpur (425405), Dist.-Dhule, Maharashtra (India).

[2]Department of Research and Development, Micro Labs Ltd, CTS No. 73, Saki Estate, Andheri East, Mumbai (400072), Maharashtra (India).

[1]nalerohini03@gmail.com, [2]chandrakantw@microlabs.in, [3]pankajs@microlabs.in, [4]divyanka.bodas@microlabs.in, [5]nidhikate@microlabs.in

Abstract- Formulation development based on artificial intelligence is an appealing approach is accelerating the division of medicinal products. Artificial intelligence (AI) is an adaptable technology with a variety of algorithms that may be used in different situations. Among the most commonly used methods of administration are dose types that are solid, which material attributes (CMAs) and processing parameters are two of the consist of granulates, powdered materials, pills, capsules, and others. While Important various elements which might impact. The product's quality during the product development process. These include rates of disintegration, chemical and physical permanence, distribution of tiny particles, and the dry powder's excellent performance. A conventional trial-and-error method of product creation, however, is time-consuming, ineffective, and challenging. Recent recognition of AI as a cutting-edge and new tool to develop pharmaceutical products has drawn a lot of emphasis. In this review, a brief summary of the use of Pharmaceutical artificially intelligent systems (AI) techniques for creating a database of appropriate amount formulae, advice the data processing for preparation, and synopsis or comparison of the concept of artificial intelligence techniques, and details on applications for artificial intelligence and investigations. Particular kinds of strong dose. Pharmaceutical applications of the sophisticated method Furthermore mentioned will include deep learning-powered visualisations of data. Applying advanced artificial intelligence (AI) helps researchers and scientists to better understand and forecast the characteristics of drug formulations, and this allows improvements in pharmaceutical company's product development techniques.

Key Words- AI Tools, Solid Dosage Form, Formulation Optimization

I. INTRODUCTION

Patients most frequently receive active pharmaceutical ingredients (APIs) via solid-state formulations using a variety of administration methodsThe most prevalent amount forms are those that are solid. type of administration between the several medication formulations available on the market](1)are Varieties of granular dosage made up including a few Interfaces and the right fillers, like stabilizers, disintegrants, antioxidants, binders, granulating agents, and so on (1). Developing solid dosage forms is often a complicated process that calls for a thorough understanding of data like pharmacokinetic/pharmacodynamic modelling (PK/PD) and physicochemical parameters. Pre-formulation, drug product development, and manufacturing are among the steps involved in the development (2).

When making a formulation, many things must be thought about, including how well it dissolves, how stable it is, how polymorphic it is, how well it works with other ingredients, how well it dissolves, how bioavailable it is, how it is manufactured, and how it can be scaled up (3). Identification for Pharmaceutical products (also known as B Grades II through intravenous can be used to show low water solubility, which is one of the most important things to keep in mind when making a new formulation (4). 90% of pharmaceuticals in research and 40% of marketed medication products are reportedly classified as weakly water-soluble (3). A small therapeutic window (5), powder that doesn't flow well (4), and chemical breakdown during production (6)are some of the other problems that come up during the formulation development process. Scientists must conduct several studies to close the information gap about the difficulties faced during formulation development. These experiments are time-consuming and difficult. Because artificial intelligence is a successful strategy that has grown in strength and adaptability in recent years, it offers a solution to this issue (7).

A method of utilizing computers to mimic human intellect is known as the application of artificial intelligence (AI). In 1956, the pair Marvin Minsky and John McCarthy organized a symposium to introduce the idea. Four processes make up a conventional AI efficiency of operation, including data gathering and preparation, modelling, simulation, testing, and deployment. Putting algorithms into practice and finding trends in data to help with decision-making is known as machine learning, a subgenre of artificial intelligence. Examples of decision-making include operational choices in healthcare (8)and risk forecasting judgments (9),(10). The approaches with layers of functionality, sometimes referred to as artificial neural networks (ANN), often represent deep learning, a branch of machine learning. Artificial neural networks (ANNs), which are based on the structure of organic neurons in human brains, are better at computing and making predictions than traditional machine learning methods. A wide range of other domains, including image classification (11), recognising objects (12), segmentation of photographs, natural language processing (13), and medical image analysis (14), have also made extensive use of deep learning.

As an alternative to the traditional approachThe creation of medications using AI has been underway used widely in the pharmaceutical industry and is regarded as a powerful and promising strategy. The AI method combines fields such such as computer science, computer vision, chemical engineering, chemistry, material science, and machine learning. The foundation known as Pharma 4.0for using cutting-edge digital methods to address certain persistent challenges in the production of pharmaceuticals (15). In a 2020 paper, Wang et al. provided a comprehensive review of "Pharma 4.0" and computational pharmaceutics. In order to accomplish this, the team examined several mathematical models, molecular models, physiology based pharmacological (PBPK) models, process simulations, and machine learning models (16). The regulatory needs, difficulties, and opportunities for the pharmaceutical sector were also summarised in this Wang et al. work.

Due to the rapid advancement of AI in the industry, the global AI market for medicines is expected to reach $1.24 billion in 2022 at a compound annual growth rate of 32.3%. More importantly, pharmaceutical corporations have invested in AI startups or established joint ventures to improve pharmaceutical goods and medical equipment (17). AI applications have already significantly increased the effectiveness of clinical trials, research, and decision-making, which benefits patients, physicians, insurers, and regulators (18) Considering numerous pharmaceutical companies operating Since AI technology companies and institutions employ AI as a tool when creating new goods, this trend is still going strong. For instance, Bayer and Merck & Co. received the ground-breaking Medical Classification from the FDA in the United States for their artificial intelligence software that aids in clinical decision-making for patients with persistent thromboembolic pulmonary hypertension. Novartis, Pfizer, and the Massachusetts Institute of Technology formed the Machine Learning for Pharmaceutical Discovery and Synthesis Consortium. (MIT) to contribute to the development of useful software for the automation of the synthesis and discovery of small molecules. AstraZeneca teamed up with Ali Health to expand the Chinese pharmaceutical company so that AI could help patients acquire the greatest drugs. Figure 1 summarises the international partnerships between the pharmaceutical and AI sectors. Exscientia, a British AI company, has more agreements with pharmaceutical corporations than any other company as of June 2022, with nine, as shown in Figure 1. Second and third place went to IKTOS and GNS Healthcare, respectively. Additionally, the establishment of multiple consortiums, such as the Machine Learning Ledger Orchestration for Drug Discovery (MLLODDY) and the Machine Learning for Pharmaceutical Discovery and Synthesis Consortium (MLDPS), has greatly facilitated the search for new medications by developing a more sophisticated.

Figure 1 shows the international ties that pharmaceutical companies have made.

Fig. 1: Relationships between AI and pharmaceutical firms to produce medical products: The recently announced partnerships between AI and pharmaceutical companies—mostly pertaining to clinical research and medication discovery—are summarised here. In terms of formulation development, especially solid dosage forms, AI and pharmaceutical businesses have few collaborations, and the great bulk of research related to the creation of solid dosage forms with AI is carried out in academic institutions. This figure's data was gathered from news releases, company reports, Securities and Exchange Commission (SEC) filings, and literature [26]. * The Supplementary Materials contain a detailed list of significant partnerships between AI and pharmaceutical businesses together with the relevant references. Previously, Merative had.

In order to customise regulatory monitoring and facilitate the enhancement of patients' lives, the U.S. FDA released the "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan" in 2021 [29]. A number of AI-based devices, including IDX-DR, OsteoDetect, Guardian Connect System, ContaCT, and FibriCheck, have received FDA approval in recent years. The U.S. FDA reviews AI-based software according to its risk classification. The majority of Class I products—like the glucose level monitor reader—have the fewest dangers. The majority of AI-based software applications are classified as Class II goods, which normally require De Novo clearance or the 501(k) method. Class III items are the most hazardous and need to go through the entire premarket approval procedure. The FDA, as a regulatory agency, also states that quality by design (QbD) and the use of computational approaches in pharmaceuticals should enhance product quality. (19)

Furthermore, the U.S. FDA has recently increased its use of quantitative modelling and methodologies (QMM), including physiologically based models (20). For the assessment of bioequivalence (BE), QMM is essential. Model-informed medication discovery and development is supported by the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) (MID3). MID3's objective is to increase regulatory review and drug development's use of similar techniques and standardise them (20). The Knowledge-aided Assessment & Structured Application (KASA) quality assessment method was introduced by the

U.S. FDA in 2019. In addition to using computers to assess quality risks and regulatory standards between facilities and applications, its objective is to develop policies and programs for risk evaluation and management (20).

Over 50% of NMEs are consistently attributed to solid dosage formulation, one of the most significant dosage forms in medicines, according to the FDA's Centre for Drug Evaluation and Research (CDER). Its numerous advantages, including stability on the shelf, patient compliance, ease of transportation, and precise dose, account for this. Giving formulation scientists a brief overview of AI methodology, including the many algorithms available, and showcasing the latest AI technologies that can be utilised to create solid dose formulations are the two main objectives of this review.

II. Frequently Utilised Sources

When conducting an AI-based investigation, a database must be obtained. Building a quality database is a must for successfully creating a suitable model for formulation development. The database itself can be configured in a number of conventional ways for modelling: (1) utilising a publicly accessible external database; (2) utilising an internal experimental database that predates the public database; or (3) utilising an experimental database created with statistical data collection and analysis tools such as the Design of Experiment (DOE) tool (21). DOE is an organised approach that aids researchers in (1) comprehending the impact of various factors on responses, (2) determining the interplay between various elements, and (3) obtaining the optimal response [35]. A summary of the open-source resources that provide information on the solid dosage formulations is given in Table 1. For the purpose of developing the preliminary dataset, the table offers a clear list of data sources for formulations, excipients, and APIs. With information on 112 million compounds, 297 million chemicals, 1.5 million bioassays, 296 million bioactivities, 185 thousand proteins, and 43 million patents from 871 organisations worldwide, PubChem is one of the most popular and widely used websites for chemical information. The U.S. FDA's inactive ingredients database has information on 9438 inactive chemicals as of August 2022 in seven categories: inactive ingredients, routes, dosage forms, CAS numbers, UNII, potency quantities, and potency units. Biological products and drugs approved for human use in the US are listed in the Drugs@FDA database. Reference listed drugs (RLD), reference standards (RS), therapeutic comparability (TE) codes, market status, producer, dosage, and other information regarding approved medications are provided by the database. One Additionally, the database includes FDA staff reviews, regulatory data, and the most recent FDA-approved labelling information for prescription brand-name drugs. Scientists can use data mining from open-source or corporate databases to produce high-quality datasets for machine learning models. The U.S. FDA's inactive ingredients database has information on 9438 inactive chemicals as of August 2022 in seven distinct groups: inactive ingredients, routes, dosage forms, CAS numbers, UNII, potency quantities, and potency units. The biological products and drugs approved for human use in the US are listed in the Drugs@FDA database. Reference listed drugs (RLD), reference standards (RS), therapeutic equivalents (TE) codes, market status, manufacturer, dosage, and other information regarding approved medications are provided by the database. Additionally, the database includes FDA staff reviews, regulatory data, and the most recent FDA-approved labelling information for prescription brand-name drugs. Scientists can use data mining from open-source or corporate databases to produce high-quality datasets for machine learning models.

Table 1: A few common databases that provide information on solid dose formulations. Numerous Admired Solid Dosage Formulation Databases.

API Chemicals	**Name**	**Size**	**Publisher**	**Reference**
Chemicals and APIs	The US Pharmacopoeia	Over 5000	Convection in the US	U.S.Pharmacopeia
	The PubChem	Over 111 million	National Centre For Biotechnology Information	(22)
	Database of Cambridge Structures	More than 900,000.	Cambridge University	Cambridge Crystallographic Data Centre
	SciFinder	142 million	Service for Chemical Abstracts (CAS)	(23)
	Index of Merck	More than 10,000	Chemistry Royal Society	Merk Index
Excipients	Search for Approved Drug Products with Inactive Ingredients	9438	FDA in the United States	FDA Approved Drug
Formulations	FDA-approved drugs, or drugs@FDA	Over 20,000	FDA in the United States	FDA Approved Drug

	Orange Book: Authorized Pharmaceutical Items with Assessments of Therapeutic Equivalence.	N/A	FDA in the United States	Orange book approved drug products
Formulations	Drug Bank	Over half a million	Alberta University	(24)
Formulations	Methods of Dissolution	1388	FDA in the United States	Dissolution Method
Formulations	MedlinePlus®	About $1500	Health National Institute	National Library of medicine
Formulations	Portal for Drug Information	more than 49,000	Health National Institute	U.S national Library of medicine

III. Approaches to Data Processing

IV. I. Processing Tabular Data

The raw database that scientists have acquired from open sources or internal experimental results must be processed before they can create the models. Some common techniques, such as data cleaning, dimension reduction, unbalanced data solutions, and data splitting, are necessary to make changes and then assess the data. One processing method for inaccurate or missing dataset observations is data cleaning. One can either remove the data elements or substitute mean/median values for them in order to do data cleaning (25). Nevertheless, there can be some disadvantages to eliminating missing values; for example, reducing the data quantity could affect the robustness of the model (26). Reducing dimensionality entails eliminating unnecessary database components. This simplifies the model and could address overfitting problems. Numerous dimensionality reduction techniques have been applied to data processing. These consist of principal component analysis (PCA), low variance filtering, high correlation filtering, and random forest feature selection (27). Unbalanced data is the term we occasionally use to describe the unequal distribution of different classifications inside a database. When there are few examples from the minority class, the overall accuracy of the model is biased towards the majority class. This indicates that the outcomes of the prediction metrics, particularly the accuracies, don't truly reflect what the In addition, certain metrics, like Cohen's Kappa and the Receiving Operating Characteristic (ROC) Area Under the Curve (AUC), are more accurate than others when evaluating a model's performance with unbalanced data (29). Data splitting is another essential data processing step. We often divide the entire data set into three subgroups at random for this process: testing, validation, and training. The data that teaches the model to make predictions is known as the training dataset. To avoid overfitting, we use test and validation subsets to validate the model. The conventional training, validation, and testing ratios are 70%/20%/10%; however, the volume of data necessitates modifications. Therefore, data processing and splitting procedures are necessary before starting computational activities.

III. II. Methods of Molecular Representation for Excipients and APIs

When working with the API of the original database and excipient molecular information, it is essential to convert the dataset into machine-readable formats. A method known as "molecular representation" (30) uses atomic configurations and chemical compositions to encode chemical identities. To create accurate molecular representations, this variational autoencoder (VAE) algorithm uses three of the most significant structural abstractions: the International Chemical Identifier (InChI) [57], the Simplified Molecular-Input Line-Entry System (SMILES) [56], and the Molfile (MDL) [31]. In addition, we employed additional representation methods in formulation development and chemical processes, such as molecular descriptors and extended-connectivity fingerprinting (ECFPs) (32, 33, 57). It is possible to implement molecular representation with some open-source applications, such as RDKit.

IV. An Overview of AI Algorithms during Creation of Granular Forms for Dosing

A range of AI-based models have been successfully employed in the creation of pharmaceutical solid dosage forms in recent years. Combining information technology, data analytics, and mathematics produces artificial intelligence. ML is a subfield of AI that is frequently divided into three types: learning through reinforcement, learning without supervision, and supervised learning (Figure 2). A sort of technique called supervised learning uses a set of input data to predict output or target variables. To achieve the necessary level of precision, the training process generates a function of the input vs. anticipated output.

Granular dose formulations have been made using a wide variety of supervised learning algorithms. K-nearest-neighbours' algorithms (KNN), Random Forest, XGBoost, LightGBM, decision trees, and Support Vector Machine are a few of them [56, 59, 60, 61]. Unsupervised learning consists of clustering and feature-finding algorithms that only control the input variables. Reinforcement learning depends on specific decisions made in a specific environment where the computer will receive rewards or penalties for its actions in order to train the model to operate at its peak efficiency. Deep learning (DL), a branch of machine learning (ML) that learns from massive amounts of experimental data, uses state-of-the-art methods such as artificial neural networks.

To make predictions, deep learning algorithms adopt an extremely intricate model structure. Multi-layered neural networks' neurones are usually used in DL models to convert data in a non-linear way. The pharmaceutical sciences have successfully employed convolutional neural networks and recurrent neural networks in recent years to create solid dosage formulations and perform a variety of practical tasks, including identifying tablet flaws (34), predicting how stable tablets will be in storage [59], predicting particle flow [36], and predicting how drugs will dissolve [67]. The Python programming language, Lisp, C++, JavaScript, and Haskell are among the programming languages used in the AI modelling process to create algorithms. Furthermore, some tools and platforms that are sold commercially, including TensorFlow, Cortana, IBM Watson, Azure Machine Learning Studio.

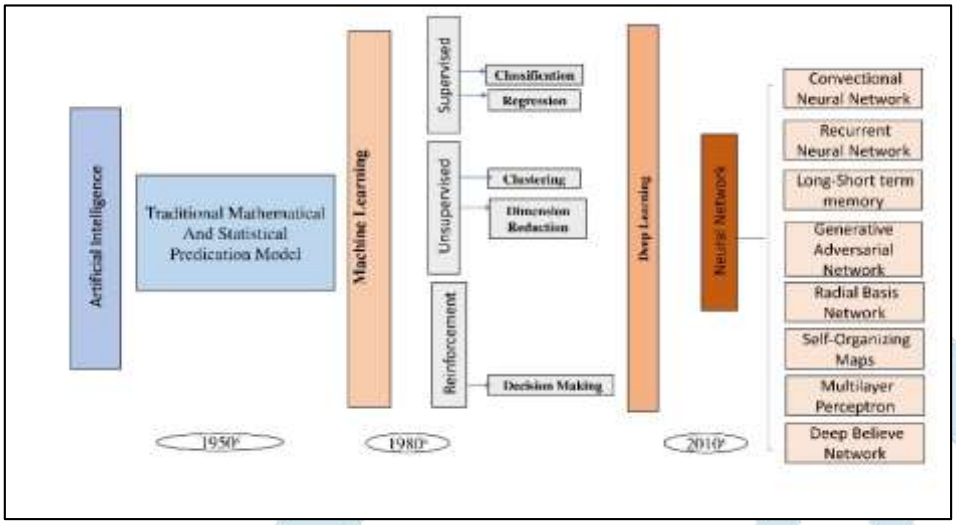

Figure 2 AI's and its subfields' development timeline.

IV. I Benefits and Drawbacks of Various Algorithms

Finding the best machine learning algorithm for the modelling process is essential when data mining and processing are finished. More importantly, certain aspects of the modelling process (such as training time, cost, and inference time) and the database (such as dimensionality, size, and complexity) need to be considered. Therefore, we have included the advantages and disadvantages of a few sample algorithms in Table 2 below. General guidelines for choosing appropriate algorithms at the outset of model building are given in Table 2. Regression analysis with independent variables is one example of a Regression using linear regression is used to predict dependent ones. The cost function and gradient descent methods help optimise the model by lowering the error. Random Forest (RF) is a popular categorisation approach. An ensemble model made up of multiple decision trees is referred to as a prediction model (RF). More importantly, a large number of incompatible trees perform better together than individual trees because they avoid errors. K-means clustering is a crucial unsupervised learning technique that groups a lot of data points based on their unique similarities. A random selection of centroids for clusters is the starting point for K-means clustering, which then iteratively maximises the centroids' positions. One popular deep learning technique in the pharmaceutical industry is the artificial neural network (ANN) (21).

An ANN is short for the input, concealed, and output layers. There are a specific number of neurones in each layer. Artificial neural networks (ANNs) use neurones to send signals and replicate the physiological processes of the human brain. The Convolutional Neural Network (CNN) is another well-liked deep learning method for image processing. A conventional CNN consists of convolutional, pooling, flattening, and hidden layers. CNNs are now widely used for image classification and segmentation applications due to the development of processing technologies, particularly GPUs and CPUs (35).

Table 2. An overview of each AI algorithm's benefits and drawbacks.

Algorithms	Advantages	Disadvantages
Regression using linear models	Simple to put into practice Effective in training It functions effectively with data that is linearly separable.	At risk for noise and overfitting assuming that the dependent and independent variables are linear
Regression Lasso	carries out variable selection and shrinkage Excellent forecasting and interpretation	The choice of models is erratic.
Ridge regression	Prevents overfitting It functions effectively when high-dimension data is available. does not call for objective estimators	Inability to select features eliminates bias for variance and lowers the coefficient to zero.
The k closest neighbours grouped	Lack of training sessions Easy to implement	Aware of outliers and missing values The system does not function well with high-dimensional

		information.
Forrest Randomly	Excellent results even with unbalanced data Reduces the number of mistakes Able to handle large datasets Effective management of missing data Reduced influence of outliers.	More easily overfitted. Poor accuracy The Black Box algorithm
Bayes's Naïve	Databases that can grow Quick and real-time forecasts Accommodating high-dimensional data	Unsatisfactory estimator performance Contrary to popular belief, not all variables are independent.
Clustering with K-means	Easy to put into practice Quick calculation time with a large number of variables Able to immediately bounce back after failure	Sensitive to outliers and noisy data The number of clusters (k) must be predetermined.
Applications'Spatial Clustering Using Noise Clustering	It is not necessary to specify the number of clusters in advance. It works nicely with clusters of any shape. Sturdy against outliers	Low performance when the dimensions of the data are large The system fails when there is fluctuating cluster density.
Clustering using mean shift	It is applicable to intricate clusters. Sturdy against outliers All that is required to calculate the number of clusters is bandwidth.	Poor performance when dealing with high-dimensional data Slow implementation time
CNN	high accuracy when CNN is optimized able to recognize patterns or significant characteristics in the pictures	increased processing power, particularly GPU, and a large amount of training data
RNN	The system can simulate a group of documents. The presumption that every pattern is reliant on its predecessors You can use convolutional layers in conjunction with it to expand the pixel neighborhood.	Disappearing gradient RNN is challenging to train. Slow computing time

V. Model Predictive Explainability and Performance Assessment

The accuracy of the models' predictions must be evaluated once the machine learning modelling process is finished. Numerous metrics, which we can divide into regression and classification metrics, can be used to assess the prediction performance. Table 3 provides a summary of the numerous metrics in different applications. Regression modelling assignments are frequently evaluated using the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). Before determining metrics like accuracy, precision, recall, F1-score, sensitivity, and specificity for classification modelling jobs, we will utilise a confusion matrix. Nevertheless, the results of certain classification metrics, such as

accuracy, are misleading in the unbalanced classification modelling challenge. As a result, In recent years, the pharmaceutical industry has seen a considerable increase in the use of deep learning-based image analysis, a subfield of machine learning, to develop solid dosage formulations (37). The evaluation criteria for comparing machine learning models for tabular data with deep learning-based image analysis are based on the number of pixels or voxels in the images. As such, other metrics such as mean precision (AP) and mean average precision (mAP) are frequently employed for object identification tasks (38). For image segmentation tasks, we employ intersection-over-union, pixel accuracy, and dice coefficients (39).

Table 3: A summary of the various metrics used to assess machine learning for categorisation and regression.

Parametric Measurements	**Factors of Division**	**Analysis of Images**
Determination coefficient (R2)	Recall, accuracy, and precision	Mean Average
MSE stands for mean squared error.	F1-score	Accuracy
RMSE, or root mean squared error	Receiving Operating Characteristics; Sensitivity and Specificity	Accuracy of Pixels and Dice Coefficient
MAE, or mean absolute error	The Kappa of Cohen	The intersection of unions

Model explainability and feature importance are specific actions to perform after the model assessment. The prediction model assigns feature relevance scores to each variable; the higher the score, the more significant the variable. To determine the meaning of the trained models, a variety of feature significance techniques were applied. These comprised major factor, maximum information coefficients, permutation importance, absolute importance, F-statistics, and contamination reduction (40, 41). The most significant features at the sample level and the reliability of each prediction have also been determined using certain sophisticated model-explainability techniques, such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). For example, Szlk et al. successfully applied SHAP to a machine-learning model of oral dissolving tablets (42). In this study, a deep neural network was trained to predict the disintegration time of tablets, yielding an R^2 of 0.84. The SHAP values of 39 input variables were computed and analysed to evaluate the model's explicability. According to the SHAP data, longer disintegration periods are frequently associated with higher disintegrant and filler concentrations. However, there was an inverse relationship with the amount of lactose. Furthermore, Ye et al. evaluated the cyclodextrin (CD) formulation model after the fact using LightGBM using the LIME approach (43). By summing together the number of hydrogen bond donors, the aromatic rings of API, the minimum projection radius of CD, and the binding-free energy during the creation of a CD complex.

VI. Applications of AI in Granular Forms of Drugs

VI. I. Summary of Granular Formulation Designs Developed by Intelligence

Because of their ease of use and patient compliance, solid dosage forms like tablets, powders, and granules continue to be the most widely used medicine product types. These goods are quite likely to continue to dominate the pharmaceutical sector in the future. Scientists and academics began examining solid dosage forms for AI applications in the 1990s. According to the published literature, the number of publications on AI in solid dosage forms has increased by 100% annually since 2015 (21). Of all the solid medication forms, tablets are the most aesthetically pleasing, and they make up over 60% of the development of solid dosage forms related to artificial intelligence (21). We provided a comprehensive overview of this research issue by compiling the most recent AI.

Table 4: An overview of the various solid dosage forms for AI applications

Dosages Forms	**Applications**	**Algorithms**	**References**
Tablet	Forecasting the release of drugs	ANN, SVM, decision tree, and ensemble of regression trees	(44),(45)
Tablet	Constructing tablets via 3D printing	CNN, SVM, RF, ANN, and self-organizing maps	(35)(37), (46)
Tablet	Recognising tablet defects	CNN's "You Only Look Once v5"	(44),(35)
Tablet	Disintegration rate estimation	CNN, ANN, XGBoost, and RF	(42),(47)
Tablet	Inspection of drug particle size	An neural network for pattern identification	(46)
Powders	Powder engineering process control	ANN	(48)
Powders	Creating an inhalable dry powder design	CNN, SVM, KNN, ANN, XGBoost, LightGBM, and RF	(49) ,(50)
Powders	Forecasting the distribution of spray-dried powder's particle sizes	Not specified	(51)
Capsule	Finding flaws in capsules	CNN, SVM, and KNN	(52)

Capsule	Finding flaws in the pellets within the capsules	SVM	(53)

VI. II. Tablets

Among the most important kinds of oral solid dose is a tablet. Tablets, typically made by compression or molding, combine APIs and excipients (54). There are three types of excipients used in tablet preparation: (1) lubricants, glidants, binders, and diluents improve the performance of tableting; (2) sweeteners and food pigments hide taste and make tablets look better; and (3) disintegrants and sustained-release polymer coating change how the drug is released. This section will cover a number of AI applications in tablet formulations, such as medication release prediction, manufacturing process optimization, and tablet fault detection.

VI. II. I. Predicting Drug Release

Among the most basic preclinical tests carried out during product development are drug release studies, which include both in vitro and in vivo research. Crucial processing parameters and material properties influence drug release patterns. For instance, little adjustments to The rate of disintegration may be significantly impacted by the compaction parameters (such as pressure and tablet form) or other factors (such as medication loading). For conventional in vitro drug release studies, certain equipment is also needed, including apparatuses, UV-visible spectroscopy instruments, and USP-approved containers. It takes a long time to analyse this traditional study in its entirety. By saving time and money, artificial intelligence (AI) technology has aided scientists in anticipating important features of medication formulations, improving the process of creating new medications. The remainder of this section will describe three published studies that predicted the drug release profiles, dissolving profiles, and disintegration times of various tablet types using artificial intelligence.

In a 2021 study, the dissolving patterns of hydrophilic matrix sustained-release tablets were predicted by three AI systems. For data analysis and dissolution profile prediction, this study used SVM, ANN, and an ensemble of regression trees. In order to build a database for modelling, we also combined the results of Process Analytical Technology (PAT) and important Material Attributes (CMAs) as input data. The results show that one of the key elements affecting the model's prediction is particle size distribution (PSD). Furthermore, the evaluation metrics demonstrated that ANN was the most accurate model among them all.

In another 2012 study, the prediction of drug release in matrix tablets was evaluated using multilayer perceptrons, decision trees, and an Elman dynamic neural network. Researchers used either polyethylene oxide polymer or glyceryl palmitostearate to generate a number of tablet matrices with different compression pressures. The CMAs were one among the input criteria, along with other tablet properties including porosity and tensile strength. By computing the difference (f1) and similarity (f2) between neural networks optimised with Monte Carlo, we evaluated the models' accuracy. As a component of RNN, the Elman dynamic neural network outperformed all the others, according to the results, and was able to predict medication release dates with high accuracy (56).

Lastly, researchers investigated the prediction of pill dissolving time using deep neural networks (57). In this study, the data on 145 medication formulations were separated into training and testing sets using the enhanced maximum dissimilarity approach (MD-FIS). Literature mining was used to gather the data. One notable illustration of an advanced data selection method is MD-FIS. A deep neural network with ten hidden layers and fifty neurones in each layer was built for the modelling process. The optimised deep neural network achieved high accuracy of 80% and 85% in the testing and validation sets, respectively. As a result, we have successfully predicted tablet medicine release characteristics using ML models.

VI. II. II AI-Powered 3D-Printed Tablet Development

Among the most cutting-edge methods for customized medicine is three-dimensional (3D) printing, which has the ability to create tablets based on a patient's physiology, genetic makeup, and medication reaction (58). To make personalized 3D-printed tablets, different methods have been used, as stereolithography, binder jetting, selective laser sintering, pressure-assisted microsyringe, and fused filament fabrication. The final product quality is significantly impacted by a number of factors, including the printing speed, platform temperature, and nozzle temperature. These variables might also impact the release of the medications both in vitro and in vivo (59). As a result, AI technologies have a lot of potential for getting incorporated into this strategy and determining the design window to enhance the 3D printing procedure and lessen the experimental load with different variables. The use of AI to improve processing parameters during the 3D printing process is illustrated in the following publications. Obeid et al. investigated the effects of processing variables and tablet surface area/volume ratio using the ANN model. First, we used self-organising maps (SOM) to visualise and interpret the interaction between different factors, and then we built a three-layer artificial neural network (ANN) for modelling, with (1) two neurones in the first layer, (2) three hidden neurones in the second layer, and (3) five neurones in the third layer. The dissolution rate was set as the goal of this study, and processing factors, such as an infill pattern and an infill density between 20% and 100%, were used as input variables. After ANN modelling and validation, we achieved high dissolution rates using a zigzag infill pattern and a smaller infill density (<50%). Most importantly, the Using 968 formulas from published literature, another recent work successfully predicted the feedstock qualities, printability, and processing temperature of 3D-printed tablets using a variety of machine learning models (RF, SVM, and ANN). Filaments for 3D printing tablets utilising fused deposition modelling (FDM) technology must be created using hot melt extrusion. The diameter, strength, and texture of the filament are all influenced by the extrusion temperature, a critical production parameter. Therefore, it is necessary to optimise the extrusion temperature in order to retain product quality.

According to this investigation, ANN accurately predicted the extrusion temperature, with a mean absolute error (MAE) of 5.18°C and an R^2 of 0.90. Additionally, the RF technique was able to predict the printing temperature, a processing parameter that impacts printability, with an R^2 of 0.86 and a mean absolute error (MAE) of 6.87°C (37).

More importantly, deep learning has demonstrated great potential in 3D-printed tablet flaw discovery by providing a computer-assisted, non-destructive quality assurance technique. To determine whether it was feasible to identify defects in tablets created using the selective laser sintering 3D printing technique, Westphal et al. employed CNN (35). They employed a variety of CNN pre-trained models for this image classification job, such as Xception and VGG16. The results showed that, at 95.8%, VGG16

had the highest accuracy. Additionally, Grad-CAM performed well in demonstrating and elucidating the CNN model, while VGG16 outperformed Xception in identifying the locations of the effects.

More importantly, deep learning has demonstrated great potential in 3D-printed tablet flaw discovery by providing a computer-assisted, non-destructive quality assurance technique. To determine whether it was feasible to identify defects in tablets created using the selective laser sintering 3D printing technique, Westphal et al. employed CNN (35). They employed a variety of CNN pre-trained models for this image classification job, such as Xception and VGG16. The results showed that, at 95.8%, VGG16 had the highest accuracy. Additionally, Grad-CAM performed well in demonstrating and elucidating the CNN model, while VGG16 outperformed Xception in identifying the locations of the effects.

VI. II. III. Finding Tablet Defects

Cracking, capping, binding, and adhering are typical characteristics that occur during the manufacturing process of tablets. Usually, a large workforce is required to manually filter out these faulty pills, which makes the process difficult to scale up. Methods like X-ray computed tomography (XRCT) can resolve this issue by analyzing the interior structure of tablets (61)(62). Researchers have effectively detected tablet faults by combining XRCT with deep learning in order to broaden the use of this technology (61).Researchers have investigated examined the detection of internal tablet flaws using convolutional neural networks. We produced multiple batches of tablets with excipients like mannitol and microcrystalline cellulose for this study, and we used XRCT to take pictures of them so we could analyse the images. The quantity of photographs rose from 573 to 43,548 with the use of an image augmentation approach. For image analysis, a CNN including three modules was utilised: (1) UNet A, which extracts the tablets from the bottle; (2) Module 2, which locates individual tablets by automated analysis; and (3) UNet B, which detects cracks inside the tablets. During the model testing, the UNet neural network showed up to 94% accuracy for seven batches of tablets. Additionally, this CNN strategy may significantly reduce time, effort.

VI. III. Powders

Some of the earliest and most common types of medicinal doses are powders. Powders consist of finely split particles in a dry material. Numerous additional dosage forms, such as pills and capsules, are based on powder which usually have particle sizes can be produced by crushing, grinding, or communicating and range in size from 10 nm to 1000 µm. Effective application of AI technology in powder engineering process control has benefitted biologics as well as small drugs. Additionally, certain studies have demonstrated the great potential of AI applications in carrier-based dry powder inhalation (49), (1).

VI. III. I. AI Applications for Powder Engineering Process Control

The powder engineering process uses micronization and other methods to make particles that are the right size for different ways of giving them, like breathing them in or eating solid dose forms When making pharmaceutical products, particle size is very important because it impacts things like surface area, porosity, solubility, powder flowability, bioavailability, and shelf life. For example, aerodynamic particle size is critical for pulmonary medicine delivery for powders that are too large (>5 µm) to reach the lung or too small (<1 µm) to be inhaled. If the particles are the proper size, a variety of methods, including jet milling, spray drying, supercritical fluid, co-crystallisation, and wet polishing, can be employed to create the required powders (63), (64), and (65). Importantly, drying temperature, pressure, air flow, and energy input are important processing variables that affect the quality or fundamental properties of the finished product. According to recent studies, it is possible to use AI technology to control crucial characteristics or product quality during the particle engineering process.

Chauhan examined the effects of different drying methods, such as freeze-drying (FD) and spray-drying (SD), on the stability and bioactivity of peptides using machine learning technology. The natural peptide network (NPN) of rice was first processed in this study utilising both SD and FD. An ANN model was then used to predict the bioactivity of peptides, including their anti-inflammatory properties. The findings demonstrated that the estimators could approximate the actual anti-inflammatory impact up to 85% of the time and that there was no discernible difference between the various drying techniques [86]. To further comprehend the drying kinetics, Keskes et al. employed AI models based on SVM and ANN. They also looked at the effects of key variables on drying time, including starting mass, drying temperature, water content, and drying pressure. Using an R^2 value.

Another study used a multilayer perceptron artificial neural network (ANN) to predict energetic performance during the SD procedure. As input variables, we employed processing variables such as drying air temperature, spray airflow rate, aspirator rate, and peristaltic pump rate. Exergy efficiency, entropy production, intake exergy, and outlet exergy were among the metrics used to quantify the energetic performance. AI may be useful in forecasting how well exergy will function in the SD process, according to the results, which revealed an R^2 of 0.98 for the ANN model.

VI. III. II. AI Use in the Design of Dried Granular Inhalation

Consumers prefer a capsule-based powder-type inhaler, and one of the most common dosage forms for administering drug formulations to the human lungs is dry powder (66). Aerosol performance is a key factor for creating a dry powder for inhalation that needs careful management. To find out how well dry powders work as aerosols [1(67)], you can use a next-generation impactor or a cascade impactor to measure the median mass aerodynamic diameter (MMAD), geometric standard deviation (GSD), and fine particle fraction (FPF). Anticipating these issues and creating dry powder inhalation products requires modelling them using AI approaches.

Using a machine learning approach, we looked at the design of a dry powder for inhalation that is based on carriers. For our research, we used sixty-five datasets, each of which had three carriers and three drugs. The database was updated with (1) CMAs and (2) quantitative variables from scanning electron microscopy (SEM) pictures. These factors included root mean square deviation (Rq), skewness of the evaluated profile (Rsk), and mean polar facet orientation (FPO). The fine particle fraction and the emitted dosage (ED) were the results. A feedforward ANN model was built for modelling after the database was divided into 50 subsets for training and 15 subsets for testing. The accuracy of the model was much higher than that of empirical modelling, with R^2 values for FPF and ED being 0.9820 and 0.9556, respectively. This piece demonstrated

VI. IV. Capsules

Drugs in capsule form are enclosed in a shell of gelatin or similar material. The capsule is another common solid dose type that works particularly well for oral administration. However, little research has been done on using AI approaches to create capsule-based formulations. The drug powders have been placed into several capsule forms, including enteric, soft gelatin, hard gelatin, and modified release capsules, to obtain various drug release patterns. Zhou et al. showed how to identify capsule problems using an enhanced CNN (52). For this study, capsules that were shrivelled, locked, or nested, as well as those with a variety of defects, including holes, concave heads, uncut bodies, and oil stains, were initially created by hand. To counteract the overfitting of the

model, we fitted an Adam optimiser and L2 regularisation to the updated CNN. K-Nearest Neighbour (KNN) and Support Vector Machines (SVM) were also used in this study as comparison tools. This enhanced CNN model was able to detect capsule faults with an accuracy of up to 97.56%, according to the confusion matrix results (52).

VI. V. Granules

Granules, a solid pharmaceutical dose formulation, comprise aggregates of powdered particles containing medications and excipients. Granules are more shelf-stable than liquid versions and provide a flexible administration option for individuals who have trouble swallowing pills or capsules. When compared to medication powder in bulk, capsules also provide better compressibility and flowability. There are several types of this dosage form, including Modified-release particles, covered pellets, glowing granulates, and gastro-resistant granules that Wet granulation and dry granulation are the two main methods of granule production (68). Recent research has demonstrated that artificial intelligence (AI) tools can be utilised to make granules, specifically to regulate the process and predict the final particle size (69).

Mariana Landin used an AI tool that employed neuro-fuzzy logic and gene expression programming to investigate the optimal impeller power for high-shear wet granulation processing. The input variables employed in this study's modelling approach included volume, wet mass density, mean torque, liquid ratio, impeller diameter, and speed. The prediction results for batches ranging in size from 25 cubic meters to 600 cubic feet showed an excellent correlation (R2 > 86.78%). Furthermore, the findings demonstrate that estimating the maximum impeller power is a practical method of determining the end of the high-shear granulation process (69).

This study was the first to demonstrate that near-infrared (NIR) spectroscopy might be used to determine the drug content of granules. The remaining medication was then predicted using a variety of machine learning algorithms based on the NIR spectra. Lastly, we improved three AI modelling development methods: backpropagation artificial neural networks (ANN), particle swarm optimisation (SVM), and a genetic algorithm. The results demonstrated that AI models are practical tools for determining how much medication is contained in granules.

In other studies, AI methods like artificial neural networks (ANN), genetic engineering, and multiple linear regression were employed to model the particle size distribution of the final granules. Using an oscillating mill, we initially made the granules with lactose, mannitol, and microcrystalline cellulose. Following that, process variables like compaction force and impeller tip speed as well as material properties like real density were used for analysis and modelling. With an NRMSE of 2.28% and R^2 = 0.9926 on the prediction scale, ANN outperformed the other two AI models in terms of prediction accuracy, per the assessment metrics data [60].

VI. VI. Solid Dispersions

Liquid fragments, one of the most crucial solubilisation methods for solid dosage forms, are often produced by polymers and pharmaceuticals. Amorphous solid dispersions are one type of solid dispersion. An amorphous drug is combined with a polymer matrix at the molecular level to create a homogeneous drug-polymer solution, also referred to as an amorphous solid dispersion. A solid dispersion can be produced via (1) solvent-based techniques like spray-drying and co-precipitation and (2) fusion-based techniques like hot-melt extrusion.

However, environmental factors including heat, moisture, storage duration, and the interaction between the drug and polymer make solid dispersion physically and chemically unstable, which eventually causes the drug and polymer to phase separate. To prevent phase separation, stability, miscibility, and solubility are critical characteristics to consider while creating a solid dispersion product. Solvent casting and other high-throughput screening techniques are feasible, but the pre-formulation, formulation, and characterisation stages of a conventional solid dispersion development process are difficult and time-consuming. Certain characteristics of solid dispersion formulations, including their dissolve rate, dissolution profile, and physical or chemical stability, have been successfully predicted by some AI-based techniques. This has been completed.

VI.VI. I. Forecasting Safety, Organic or Physiological

The study examined the feasibility of forecasting the physical stability of solid dispersion using a variety of machine learning algorithms, including ANN, SVM, RF, LightGBM, KNN, and naive Bayes [59]. In order to train the model, 646 physical stability data points—including 50 pharmaceutical compounds—were gathered from a public source. Molecular representations of attributes such as molecular weight, melting temperature, hydrogen bond acceptor count, and heavy atom count were used in the creation of the database. An accelerated stability study was conducted for three and six months at 40°C and 75% RH to evaluate the model's ability to predict physical stability. With an accuracy of 82%, RF was the most accurate forecasting model. Next, we made a solid dispersion of 17-oestradiol via solvent evaporation. Volume, temperature, and drug loading were among the processing characteristics that were combined with the chemical descriptors as input variables. Another output and regression objective were the disintegration rate. In contrast, the dissolution type was a binary output that could result in either precipitation or supersaturation. With an accuracy of up to 97.7% for dissolution type, we found that ECFP4-XGBoost was the most accurate. Random Forest, SVM, and LightGBM all exhibit exceptional accuracy in forecasting the dissolution rate (R^2 ranges from 0.809 to 0.928).

Experimental validation of machine learning models is also crucial to confirm the model's predictive power and conformity to real-world situations. In their investigation, Gao et al. meticulously planned and produced a ternary solid dispersion using computer-aided technology. The binding-free energy of the interaction between the medication andrographolide and various forms of cyclodextrin was estimated in the first section of this study using a LightGBM model. Then, a study of its solubility supported the theory that γ-cyclodextrin had the strongest binding affinity. To further comprehend the inclusion process, we employed molecular dynamic modelling. Tests on both cells and animals demonstrated the effectiveness of the ML model. By adding D-α-tocopherol polyethylene glycol succinate (TPGS), the absorption into cells was significantly enhanced. In general, the model worked.

VI. VII. Using AI to Analyse Pharmacological Information

Visual examination is one of the most crucial methods in pharmaceutical research for determining the types of APIs, excipients, and dosage forms as well as their quantities [123]. Visual screening for injections must guarantee that the final products are free of invisible particles, according to the US Pharmacopoeia (USP). This kind of image analysis was condensed into static and

dynamic picture analysis by Farkas et al. [123]. In static image analysis, the particles stay still while the image is being taken. Among the techniques that could be employed to capture the images are bright field microscopy, confocal laser scanning microscopy, fluorescence microscopy, microfocus X-ray imaging, scanning electron microscopy, and polarised light microscopy.

Accurate results are also obtained by dynamic image analysis, which makes use of instruments such as process analytical technology (PAT), in-line photometric stereo imaging, and dynamic foam analysers. Over the past few years, deep learning has been widely used in medical image analysis, which has helped with tasks including picture classification, object recognition, picture segmentation, registration, and more. Additionally, the deep learning-based image analysis technique has lately been applied in pharmaceutical research and has shown considerable potential in certain fields. The next paragraphs will specifically discuss three published research studies that employed AI to measure particle size, estimate drug release in vitro, and determine how quickly tablets break apart. We believe that deep learning-based picture analysis provides significant benefits for creating solid dosage formulations, such as high accuracy, high

VI. VII. I. Image Pre-Processing Methods

Deep learning-based image analysis has garnered more attention in the pharmaceutical industry since it is crucial to ensure image quality before supplying images to the models. Therefore, some image pre-processing techniques, like scaling, normalisation, contrast correction, and image enhancement, must be used in order to create a powerful model. Batches of photos should have uniform widths and heights before being loaded into the model. Additionally, all images should have the same aspect ratio, or the proportion of width to length. Brightness and contrast are other essential components for image pre-processing. Normalisation is the process of rescaling the pixel values to 0–1, which expedites the model training process. Some images taken during trials had extremely low contrast and intensity.

Accurate results are also obtained by dynamic image analysis, which makes use of instruments such as process analytical technology (PAT), in-line photometric stereo imaging, and dynamic foam analysers. Over the past few years, deep learning has been widely used in medical image analysis, which has helped with tasks including picture classification, object recognition, picture segmentation, registration, and more. Additionally, the deep learning-based image analysis technique has lately been applied in pharmaceutical research and has shown considerable potential in certain fields. The next paragraphs will specifically discuss three published research studies that employed AI to measure particle size, estimate drug release in vitro, and determine how quickly tablets break apart. We believe that deep learning-based picture analysis provides significant benefits for creating solid dosage formulations, such as high accuracy, high.

VI. VII. II. AI-Based Imagery Processing Investigations

Successfully developed a method for measuring the size distribution of spray-dried particles using XRCT images (70). This study used an artificial intelligence-assisted approach to quantitatively analyse thousands of individual particles. The results of this image analysis have demonstrated a high correlation with the measurement of laser diffraction. More importantly, this method could facilitate the production of high-performance spray-dried nanomaterials (70).

In a separate work, Liu et al. efficiently visualised and estimated the drug release rate of long-acting parenteral implants using deep learning-based XRCT image processing. To aid in visualising the drug distribution with the implants and facilitate the interpretation of quantitative structural data, they especially employed picture segmentation and analytics. With a two-fold overestimate, the drug release prediction based on voxel computing utilising XRCT images performed poorly. The experimental and expected outcomes for the in vitro drug release were extremely similar when FIB-SEM (focused ion beam scanning electron microscopy) was employed. The changes were around 10% and 5% for implants with low and medium drug loading, respectively. This study has shown how to analyse implant growth microstructure using image analysis.

Disintegration testing is one of the most crucial stages in the development of immediate-release tablet products. However, the United States Pharmacies of <701>'s current guidance gives information on the time required for the breakdown process, which is arbitrary, liable to change, and susceptible to error by people. With an accuracy of up to 99.6%, researchers have recently created a Computer Vision for Disintegration (CVD) system that can identify and quantify the pace at which tablets disintegrate. In short, a camera took pictures of the tablet breaking apart, and a CNN was used to process and decipher the information. Beyond only examining the length, this technological platform enables a more thorough study of the tablet breakdown process.

Furthermore, this research used machine learning modeling with UV and Vis PAT imaging technologies to examine the distribution of particle sizes in tablets. This study used UV and Vis imaging along with pattern recognition neural networks to create machine vision systems that can sort particles by size. So, the refined After examining the particle size distribution of meloxicam tablets, the deep learning model achieved an excellent precision of 97%. Above all, this method might offer a rapid, non-invasive, in-line tool for characterising tablet particle dimensions.

VII. AI's drawbacks and restrictions in the development of drugs

Despite AI's advantages for drug research, there are a number of limitations and challenges to take into account. One of the biggest challenges is locating pertinent data. AI-based approaches typically require large data sets for training. In many instances, the limited amount of available data or inconsistent or low-quality data may affect the reliability and accuracy of the results. Since AI-based techniques may raise concerns about bias and fairness (see the following section), ethical issues present an additional obstacle. For example, if the data used to train an ML system is biased or unrepresentative, the algorithm's predictions may be unfair or inaccurate. Making sure AI is applied fairly and morally Another approach to attempting to clarify and simplify the reasoning behind machine learning algorithms' predictions is to use explainable AI (XAI) approaches. This could assist in addressing concerns regarding bias and fairness in AI-based systems and shed light on the underlying assumptions and methodologies that underlie the forecasts.

Current artificial intelligence (AI)-based methods cannot replace traditional experimental procedures or the expertise and experience of human researchers [50, 51]. AI is limited to making predictions based on available data; human researchers are then required to confirm and interpret the results [52]. Nonetheless, integrating AI with traditional experimental methods may enhance the drug discovery process [3]. By combining the predictive power of AI with the expertise and experience of human researchers [53], it is possible to expedite the drug discovery process and accelerate the development of new medications [54].

VIII. Prospects

Despite the widespread use of AI-based models in formulation creation, there are still certain uncharted areas that need investigation. For instance, the fields of chemistry and material science have made extensive use of sophisticated deep learning techniques like Graphical convolutional network designs (GCNs) and generative adversarial networks (GANs). Kojima and associates looked into how GCN could be used to predict how compounds and proteins would interact. The model showed how the atoms contributed to the prediction. GAN is also becoming more important in drug development because it makes it easier to explore and optimize the chemical design space for the desired function However, there hasn't been much written about the applications of GCN for short and Gac in pharmaceutical formulation development; this area might use more research. Additionally, several applications of AI/ML in solid dosage formulation processes remain unexplored. Particle size distribution is one of the numerous elements of the formulations that can be extracted by deep learning-based image analysis algorithms. To verify specific attributes in a computer simulation throughout the production process, it would be intriguing to integrate picture analysis with a process analytical technology (PAT) tool. Finally, while reinforcement learning is a potent learning algorithm that has not yet been fully explored, supervised learning accounts for the majority of the methods employed in the published literature.

IX. Conclusions

We wrapped up by talking about the usage of AI techniques to develop solid dose compositions and apply them to other formulations. While most research has proven that AI can revolutionise the drug discovery pipeline, it has also shown growing potential in formulation development. An AI-based development approach tends to expedite the process by enabling scientists to produce low-cost forecasts more rapidly. This contrasts with the conventional formulation development approach, which is labour-intensive and involves experimentation and failure. This evaluation also addressed a range of AI algorithms used for various tasks and provided general guidelines for model selection to help scientists better integrate AI technologies into their research. To offer a mechanism for systematically.

REFERENCES

1. Jiang J, Ma X, Ouyang D, Williams RO. Emerging Artificial Intelligence (AI) Technologies Used in the Development of Solid Dosage Forms. Pharmaceutics. 2022 Oct 22;14(11):2257.

2. Chow K, Tong HHY, Lum S, Chow AHL. Engineering of Pharmaceutical Materials: An Industrial Perspective. J Pharm Sci. 2008 Aug;97(8):2855–77.

3. Loftsson T, Brewster ME. Pharmaceutical applications of cyclodextrins: basic science and product development. Journal of Pharmacy and Pharmacology. 2010 Nov 1;62(11):1607–21.

4. Li J, Wu Y. Lubricants in Pharmaceutical Solid Dosage Forms. Lubricants. 2014 Feb 25;2(1):21–43.

5. Benet LZ, Goyan JE. Bioequivalence and Narrow Therapeutic Index Drugs. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy. 1995 Jul 8;15(4):433–40.

6. Hengsawas Surasarang S, Keen JM, Huang S, Zhang F, McGinity JW, Williams RO. Hot melt extrusion versus spray drying: hot melt extrusion degrades albendazole. Drug Dev Ind Pharm. 2017 May 4;43(5):797–811.

7. Bannigan P, Aldeghi M, Bao Z, Häse F, Aspuru-Guzik A, Allen C. Machine learning directed drug formulation development. Adv Drug Deliv Rev. 2021 Aug;175:113806.

8. Abdillah AA, Azwardi A, Permana S, Susanto I, Zainuri F, Arifin S. Performance evaluation of linear discriminant analysis and support vector machines to classify cesarean section. Eastern-European Journal of Enterprise Technologies. 2021 Oct 31;5(2 (113)):37–43.

9. Berk R. An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. J Exp Criminol. 2017 Jun 8;13(2):193–216.

10. Berk RA, Sorenson SB, Barnes G. Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions. J Empir Leg Stud. 2016 Mar 8;13(1):94–115.

11. Affonso C, Rossi ALD, Vieira FHA, de Carvalho ACP de LF. Deep learning for biological image classification. Expert Syst Appl. 2017 Nov;85:114–22.

12. Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, et al. Deep Learning for Generic Object Detection: A Survey. Int J Comput Vis. 2020 Feb 31;128(2):261–318.

13. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28;521(7553):436–44.

14. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep Learning in Medical Image Analysis. In 2020. p. 3–21.

15. Steinwandter V, Borchert D, Herwig C. Data science tools and applications on the way to Pharma 4.0. Drug Discov Today. 2019 Sep;24(9):1795–805.

16. Wang W, Ye Z, Gao H, Ouyang D. Computational pharmaceutics - A new paradigm of drug delivery. Journal of Controlled Release. 2021 Oct;338:119–36.

17. Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects. Drug Discov Today. 2019 Mar;24(3):773–80.

18. Jarrahi MH. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Bus Horiz. 2018 Jul;61(4):577–86.

19. 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 Jun 11;18(6):463–77.

20. Marshall S, Madabushi R, Manolis E, Krudys K, Staab A, Dykstra K, et al. Model-Informed Drug Discovery and Development: Current Industry Good Practice and Regulatory Expectations and Future Perspectives. CPT Pharmacometrics Syst Pharmacol. 2019 Feb;8(2):87–96.

21. Lou H, Lian B, Hageman MJ. Applications of Machine Learning in Solid Oral Dosage Form Development. J Pharm Sci. 2021 Sep;110(9):3150–65.

22. Kim S. Getting the most out of PubChem for virtual screening. Expert Opin Drug Discov. 2016 Sep 5;11(9):843–55.

23. Gabrielson SW. SciFinder. Journal of the Medical Library Association. 2018 Oct 4;106(4).

24. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018 Jan 4;46(D1):D1074–82.

25. Liu H, Shah S, Jiang W. On-line outlier detection and data cleaning. Comput Chem Eng. 2004 Aug;28(9):1635–47.

26. Zhu J, Ge Z, Song Z, Gao F. Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data. Annu Rev Control. 2018;46:107–33.

27. Palo HK, Sahoo S, Subudhi AK. Dimensionality Reduction Techniques: Principles, Benefits, and Limitations. In: Data Analytics in Bioinformatics. Wiley; 2021. p. 77–107.

28. Lee H, Kim J, Kim S, Yoo J, Choi GJ, Jeong YS. Deep Learning-Based Prediction of Physical Stability considering Class Imbalance for Amorphous Solid Dispersions. J Chem. 2022 Mar 18;2022:1–11.

29. Jeni LA, Cohn JF, De La Torre F. Facing Imbalanced Data--Recommendations for the Use of Performance Metrics. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. IEEE; 2013. p. 245–51.

30. Raghunathan S, Priyakumar UD. Molecular representations for machine learning applications in chemistry. Int J Quantum Chem. 2022 Apr 5;122(7).

31. Wigh DS, Goodman JM, Lapkin AA. A review of molecular representation in the age of machine learning. WIREs Computational Molecular Science. 2022 Sep 18;12(5).

32. Dong J, Gao H, Ouyang D. PharmSD: A novel AI-based computational platform for solid dispersion formulation design. Int J Pharm. 2021 Jul;604:120705.

33. Yang Q, Liu Y, Cheng J, Li Y, Liu S, Duan Y, et al. An Ensemble Structure and Physicochemical (SPOC) Descriptor for Machine-Learning Prediction of Chemical Reaction and Molecular Properties. ChemPhysChem. 2022 Jul 19;23(14).

34. Ma X, Kittikunakorn N, Sorman B, Xi H, Chen A, Marsh M, et al. Application of Deep Learning Convolutional Neural Networks for Internal Tablet Defect Detection: High Accuracy, Throughput, and Adaptability. J Pharm Sci. 2020 Apr;109(4):1547–57.

35. Westphal E, Seitz H. A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks. Addit Manuf. 2021 May;41:101965.

36. Hesse R, Krull F, Antonyuk S. Prediction of random packing density and flowability for non-spherical particles by deep convolutional neural networks and Discrete Element Method simulations. Powder Technol. 2021 Nov;393:559–81.

37. Muñiz Castro B, Elbadawi M, Ong JJ, Pollard T, Song Z, Gaisford S, et al. Machine learning predicts 3D printing performance of over 900 drug delivery systems. Journal of Controlled Release. 2021 Sep;337:530–45.

38. Padilla R, Passos WL, Dias TLB, Netto SL, da Silva EAB. A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. Electronics (Basel). 2021 Jan 25;10(3):279.

39. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging. 2015 Dec 12;15(1):29.

40. Casalicchio G, Molnar C, Bischl B. Visualizing the Feature Importance for Black Box Models. In 2019. p. 655–70.

41. Huynh-Thu VA, Saeys Y, Wehenkel L, Geurts P. Statistical interpretation of machine learning-based feature importance scores for biomarker discovery. Bioinformatics. 2012 Jul 1;28(13):1766–74.

42. Szlęk J, Khalid MH, Pacławski A, Czub N, Mendyk A. Puzzle out Machine Learning Model-Explaining Disintegration Process in ODTs. Pharmaceutics. 2022 Apr 13;14(4):859.

43. Ye Z, Yang W, Yang Y, Ouyang D. Interpretable machine learning methods for in vitro pharmaceutical formulation development. Food Front. 2021 Jun 5;2(2):195–207.

44. Ficzere M, Mészáros LA, Kállai-Szabó N, Kovács A, Antal I, Nagy ZK, et al. Real-time coating thickness measurement and defect recognition of film coated tablets with machine vision and deep learning. Int J Pharm. 2022 Jul;623:121957.

45. Salem S, Byrn SR, Smith DT, Gurvich VJ, Hoag SW, Zhang F, et al. Impact assessment of the variables affecting the drug release and extraction of polyethylene oxide based tablets. J Drug Deliv Sci Technol. 2022 May;71:103337.

46. Mészáros LA, Farkas A, Madarász L, Bicsár R, Galata DL, Nagy B, et al. UV/VIS imaging-based PAT tool for drug particle size inspection in intact tablets supported by pattern recognition neural networks. Int J Pharm. 2022 May;620:121773.

47. Floryanzia S, Ramesh P, Mills M, Kulkarni S, Chen G, Shah P, et al. Disintegration testing augmented by computer Vision technology. Int J Pharm. 2022 May;619:121668.

48. Chauhan S, O'Callaghan S, Wall A, Pawlak T, Doyle B, Adelfio A, et al. Using Peptidomics and Machine Learning to Assess Effects of Drying Processes on the Peptide Profile within a Functional Ingredient. Processes. 2021 Feb 26;9(3):425.

49. Farizhandi AAK, Alishiri M, Lau R. Machine learning approach for carrier surface design in carrier-based dry powder inhalation. Comput Chem Eng. 2021 Aug;151:107367.

50. Jiang J, Peng HH, Yang Z, Ma X, Sahakijpijarn S, Moon C, et al. The applications of Machine learning (ML) in designing dry powder for inhalation by using thin-film-freezing technology. Int J Pharm. 2022 Oct;626:122179.

51. Xi H, Zhu A, Klinzing GR, Zhou L, Zhang S, Gmitter AJ, et al. Characterization of Spray Dried Particles Through Microstructural Imaging. J Pharm Sci. 2020 Nov;109(11):3404–12.

52. Zhou J, He J, Li G, Liu Y. Identifying Capsule Defect Based on an Improved Convolutional Neural Network. Shock and Vibration. 2020 Jul 11;2020:1–9.

53. Doerr FJS, Florence AJ. A micro-XRT image analysis and machine learning methodology for the characterisation of multi-particulate capsule formulations. Int J Pharm X. 2020 Dec;2:100041.

54. Ghourichay MP, Kiaie SH, Nokhodchi A, Javadzadeh Y. Formulation and Quality Control of Orally Disintegrating Tablets (ODTs): Recent Advances and Perspectives. Biomed Res Int. 2021 Jan 24;2021(1).

55. Galata DL, Könyves Z, Nagy B, Novák M, Mészáros LA, Szabó E, et al. Real-time release testing of dissolution based on surrogate models developed by machine learning algorithms using NIR spectra, compression force and particle size distribution as input data. Int J Pharm. 2021 Mar;597:120338.

56. Petrović J, Ibrić S, Betz G, Đurić Z. Optimization of matrix tablets controlled drug release using Elman dynamic neural networks and decision trees. Int J Pharm. 2012 May;428(1–2):57–67.

57. Han R, Yang Y, Li X, Ouyang D. Predicting oral disintegrating tablet formulations by neural network techniques. Asian J Pharm Sci. 2018 Jul;13(4):336–42.

58. Vaz VM, Kumar L. 3D Printing as a Promising Tool in Personalized Medicine. AAPS PharmSciTech. 2021 Jan 17;22(1):49.

59. Alhijjaj M, Nasereddin J, Belton P, Qi S. Impact of Processing Parameters on the Quality of Pharmaceutical Solid Dosage Forms Produced by Fused Deposition Modeling (FDM). Pharmaceutics. 2019 Nov 27;11(12):633.

60. Obeid S, Madžarević M, Krkobabić M, Ibrić S. Predicting drug release from diazepam FDM printed tablets using deep learning approach: Influence of process parameters and tablet surface/volume ratio. Int J Pharm. 2021 May;601:120507.

61. Ma X, Kittikunakorn N, Sorman B, Xi H, Chen A, Marsh M, et al. Application of Deep Learning Convolutional Neural Networks for Internal Tablet Defect Detection: High Accuracy, Throughput, and Adaptability. J Pharm Sci. 2020 Apr;109(4):1547–57.

62. Yost E, Chalus P, Zhang S, Peter S, Narang AS. Quantitative X-Ray Microcomputed Tomography Assessment of Internal Tablet Defects. J Pharm Sci. 2019 May;108(5):1818–30.

63. Giry K, Péan JM, Giraud L, Marsas S, Rolland H, Wüthrich P. Drug/lactose co-micronization by jet milling to improve aerosolization properties of a powder for inhalation. Int J Pharm. 2006 Sep;321(1–2):162–6.

64. Okamoto H, Danjo K. Application of supercritical fluid to preparation of powders of high-molecular weight drugs for inhalation. Adv Drug Deliv Rev. 2008 Feb;60(3):433–46.

65. Keskes S, Hanini S, Hentabli M, Laidi M. Artificial Intelligence and Mathematical Modelling of the Drying Kinetics of Pharmaceutical Powders. Kemija u industriji. 2020;69(3–4):137–52.

66. Lavorini F, Pistolesi M, Usmani OS. Recent advances in capsule-based dry powder inhaler technology. Multidiscip Respir Med. 2017 Dec 22;12(1):11.

67. Mitchell JP, Nagel MW, Wiersema KJ, Doyle CC. Aerodynamic particle size analysis of aerosols from pressurized metered-dose inhalers: Comparison of andersen 8-stage cascade impactor, next generation pharmaceutical impactor, and model 3321 aerodynamic particle sizer aerosol spectrometer. AAPS PharmSciTech. 2003 Dec;4(4):425–33.

68. Mitchell JP, Nagel MW, Wiersema KJ, Doyle CC. Aerodynamic particle size analysis of aerosols from pressurized metered-dose inhalers: Comparison of andersen 8-stage cascade impactor, next generation pharmaceutical impactor, and model 3321 aerodynamic particle sizer aerosol spectrometer. AAPS PharmSciTech. 2003 Dec;4(4):425–33.

69. Landin M. Artificial Intelligence Tools for Scaling Up of High Shear Wet Granulation Process. J Pharm Sci. 2017 Jan;106(1):273–7.

70. Xi, H.; Zhu, A.; Klinzing, G.R.; Zhou, L.; Zhang, S.; Gmitter, A.J.; Ploeger, K.; Sundararajan, P.; Mahjour, M.; Xu, W. Characterization of Spray Dried Particles Through Microstructural Imaging. J. Pharm. Sci. 2020, 109, 3404–3412. https://doi.org/10.1016/J.XPHS.2020.07.032

71. Liu, Z.; Li, L.; Zhang, S.; Lomeo, J.; Zhu, A.; Chen, J.; Barrett, S.; Koynov, A.; Forster, S.; Wuelfing, P.; et al. Correlative ImageBased Release Prediction and 3D Microstructure Characterization for a Long-Acting Parenteral Implant. Pharm. Res. 2021, 38, 1915–1929. https://doi.org/10.1007/S11095-021-03145-2/FIGURES/12