Intriguing of pharmaceutical product development processes with the help of artificial intelligence and deep/machine learning or artificial neural network

Naitik Jariwala  
National Institute of Pharmaceutical Education and Research - Guwahati, Sila Katamur (Halugurisuk)

Chandra Lekha Putta  
National Institute of Pharmaceutical Education and Research - Guwahati, Sila Katamur (Halugurisuk)

Ketki Gatade  
National Institute of Pharmaceutical Education and Research - Guwahati, Sila Katamur (Halugurisuk)

Manasi Umarji  
National Institute of Pharmaceutical Education and Research - Guwahati, Sila Katamur (Halugurisuk)

Syed Nazrin Ruhina Rahman  
National Institute of Pharmaceutical Education and Research - Guwahati, Sila Katamur (Halugurisuk)

Datta Maroti Pawde  
National Institute of Pharmaceutical Education and Research - Guwahati, Sila Katamur (Halugurisuk)

Amoolya Sree  
National Institute of Pharmaceutical Education and Research - Guwahati, Sila Katamur (Halugurisuk)

Atul Sayaji Kamble  
National Institute of Pharmaceutical Education and Research - Guwahati, Sila Katamur (Halugurisuk)

Abhinab Goswami  
National Institute of Pharmaceutical Education and Research - Guwahati, Sila Katamur (Halugurisuk)

Payel Chakraborty  
National Institute of Pharmaceutical Education and Research - Guwahati, Sila Katamur (Halugurisuk)

Tamilvanan Shunmugaperumal (✉ tamilvanan1@yahoo.co.in )  
National Institute of Pharmaceutical Education and Research - Guwahati, Sila Katamur (Halugurisuk)

Research Article

Keywords: Artificial neural network, artificial intelligence, deep learning, machine learning, linear and non-linear models

Posted Date: October 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2112595/v1
Abstract

The objectives of current review are (1) to provide a historical overview of artificial intelligence and deep/machine learning (AI & D/ML) or Artificial Neural Network (ANN) (2) to update the financial dealings of pharma companies related to the application of AI & D/ML or ANN in drug discovery and development processes and (3) to showcase the application of AI & D/ML or ANN concept for optimization of analytical method conditions and formula of the dosage form. The optimization of analytical method conditions and formula of dosage form started with the employment of linear model such as design of experiment followed by non-linear model like AI & D/ML or ANN. Such type of linear and non-linear models blending in optimization processes nevertheless helped to suitably identify the influence of critical process parameters or critical material attributes on critical quality attributes. However, much of integration and understandable interpretation between the available data arised from clinical trials and the prevalence/progression of pandemic/endemic infections could potentially be ambitioned through the application of AI & D/ML or ANN.

Introduction

Producing pharmaceutical products without untoward effects but with desired qualities is not only the very basic requirement set by regulatory authorities but also indirectly decides the success of the pharma industry. The pharmaceutical product development process is a combined and coordinated work from numerous divisions of the pharma industry which usually starts from API discovery, synthesis and utilization to end with formulation development, market positioning and successful use at the consumer end.

Conventional API discovery and development process as well as optimization of analytical method and formula of dosage form routinely use the quality by testing (QbT) approach or one factor at a time (OFAT) strategy. Since the time consuming and chemical wastage in OFAT strategy or QbT approach is inevitable, the looking for a better alternative approach or strategy which shows less time consumption in conjunction with minimal chemical wastage becomes an urgent requirement. In terms of the response surface methodology (RSM) approach, the design of experiment (DoE) is being utilized effectively to optimize APIs synthetic process, analytical method development and formula optimization for final pharmaceutical products (Rahman et al., 2021; Rahman et al., 2020). It should be added that the RSM-linked DoEis based on a linear model and therefore it won’t consider the non-linear modelization concept to see the influence of independent factors on the response variables. Thus, depending solely on the DoEfor such optimization processes may end with an erroneous conclusion and thus necessitates applying another approach (preferably of non-linear model based) which judiciously eliminates the conclusion errors noticed with the DoE-based optimization process. One such non-linear model-based approach recently introduced in pharmaceutical product development process is artificial intelligence and deep/machine learning (AI & D/ML) or artificial neural network (ANN). Since the information processing capacity of ANN is related to the functioning of the normal human brain, the estimation of the process parameters is being carried out on the number of trials performed by varying the composition of the
excipients and processing conditions (Ghate et al., 2019). AI is a division of computer science, involved in problem resolution and in creating machines that can perform tasks which would otherwise require intelligence and human operators (Senturaman, 2020). In simple words, AI & D/ML is a branch of computer science that deals with problem-solving through the aid of symbolic programming (Krishnaveni et al. 2019). The DL consists of a neural network of multiple layers that aim to emulate how the information processing is carried out by the human brain after understanding complicated patterns and feature interactions. In practical situations, deep learning helps to give structure to unstructured data and enables machines to learn to classify data without assistance (Abhinav & Subrahmanyam, 2019; Gilvary et al. 2019). The ML is a subset of AI utilizing algorithm models and uses statistical methods with the ability to learn with or without being explicitly programmed (Abhinav & Subrahmanyam, 2019). It should be added that the AI & D/ML or ANN is based on a non-linear concept to study the independent factor’s influence on response variables. Non-linearity refers to a massive parallel network distributed throughout that allows for approximation and real-time operation to exhibit unpredictability and random behaviour. Figure 1 shows the possible areas of AI & D/ML or ANN in the pharmacy field of the healthcare system.

[Insert Fig. 1 here]

There have been several excellent articles in both reviews in general (Cavasotto & Di Filippo, 2021; Chaki et al., 2020; Damiati et al., 2020; Paul et al. 2021, Zhou et al. 2020) and an entire issue of a journal devoted solely to the subject of AI & D/ML or ANN (Artificial Intelligence). Furthermore, as one of the non-linear models for pharmaceutical product optimization, the AI & D/ML or ANN is either combined with a linear model like RSM (Ghate et al., 2019; El Menshawe et al., 2014; Naveen et al., 2020) or only ANN-supported formula optimization of dosage form (Moghaddam et al., 2010; Manda et al., 2019; Lefnaoui et al., 2020) and food products (Rakhavan et al., 2016; Dash & Das 2021; Samson et al., 2016). Similarly, the optimization of analytical method conditions for APIs are also articulated through the AI & D/ML or ANN (Arabzadeh et al., 2019) with support from RSM (Rahman et al., 2021). The objectives of this “meta-review” are (1) to provide a historical overview of AI & D/ML or ANN, (2) to update the financial dealings of pharma companies related to the application of AI & D/ML or ANN in drug discovery and development processes and (3) to showcase the application of in AI & D/ML or ANN concept for optimization of analytical method conditions and formula of the dosage form.

**Historical overview on AI & D/ML or ANN**

Table 1 delineates the historical overview on AI & D/ML or ANN in ascending order of milestone years. The ANN has a history that dates to the precomputer era although the original goal of AI & D/ML or ANN is to solve problems mostly related to biology, the extrapolation of its applicability is much more in recent years and even the intriguing nature of AI & D/ML or ANN especially in medical diagnosis principles as well as in forecasting/predicting the disease pattern in a particular region or whole world. Because firstly the AI & D/ML or ANN helps us understand the impact of increasing/decreasing disease progression vertically or horizontally on computational time. Secondly, the AI & D/ML or ANN helps us understand the situations or cases where the model fits best. Thirdly, it also explains why the certain model works better
in certain environment or situation. Samson et al., (2016) described the ANN as an information-processing paradigm that is related to biological nervous systems i.e., the human brain. In the early 1940s (McCulloch & Pitts, 1943), researchers developed the “threshold logic” model, which encompassed a two-pronged approach to computational models of ANN. This model is directed at identifying the biological neural networks separate from ascertaining the correlation of the neural networks to artificial intelligence. An unsupervised learning model utilizing neural plasticity and long-term potentiation has been introduced in the form of Hebbian learning (Morris, 1999). This type of learning was utilized in calculators and other computational instruments (Rochester et al., 1956; Farely & Clark, 1954). Development of the perception in the model added another dimension to AI & D/ML or ANN by incorporating a two-layer computer network in pattern recognition through an algorithm (Rosenblatt, 1958). Particularly, the ANNs comprise a set of nodes, each of which receives a separate input, which is finally converted to output. After the introduction of a clear definition of backpropagation (Werbos, 1975), the supervised learning method in which the ANN receives training in conjunction with optimization becomes easier to determine the loss of function during the anticipated output for each input value is known. In this way, the ANNs are linked to single or multiple algorithms to solve the problems (Paul et al., 2021). Although the AI & D/ML were coined around in the 1950s, it now becomes a slogan in pharma industries especially after finding out their benefits in handling increased volumes of raw data following the introduction of advanced algorithms (Sharma, 2019).

[Insert Table 1 here]

**Pharma companies’ financial dealings related to the application of AI & D/ML or ANN in drug discovery and development processes**

Table 2 displays the financial dealings of Pharma companies concerning the application part of AI & D/ML or ANN, particularly in drug discovery and development processes. The entry of AI & D/ML or ANN helps to shorten not only the new drug development period but also it significantly minimizes the utilization of manpower and considerable reduction in expenditure related to the API development. For example, the German-based biotechnology company, Evotec, has partnered with a UK-based company, Exscientia, for the small molecule drug discovery process. Within a short period of 8 months, the discovered small drug molecule entered Phase 1 clinical trials which might usually have taken 4-5 years to deliver the drug candidate from the traditional drug discovery process (without utilizing AI & D/ML or ANN).

[Insert Table 2 here]

**AI & D/ML or ANN in optimization of analytical method conditions and formula of dosage form**

Before entering the discussion related to optimization of analytical method conditions and formula of dosage form, it needs to be emphasized that the ANN simply mimics the principles of information processing handled by the human brain wherein the influence of critical material attributes variation on critical analytical attributes (CAAs) can be predicted by segregating different sets of data (generated from
numerous trails) into training, testing and validating (Samson et al., 2016). For this purpose, the ANN must be coupled with an algorithm to attain the “best fit” optimum values for a method (Ghaaheri et al., 2015). The ANN-linked algorithm produces a highly reliable and better predictor of the optimum values for a method than the RSM-based linear model (Sha & Edwards, 2007). However, the ANN relies on the number of experiments/trials conducted and consequently, it is highly likely that too high/a smaller number of trials would result in error and fault in the predictions (Ghate et al., 2019). Therefore, the ANN takes the trials of the linear model (RSM following face-centred central composite design (CCD) for generating the non-linear model [ANN-linked Levenberg-Marquardt (LM) algorithm] to predict the optimum regions for the studied CAAs (Rahman et al., 2021). Furthermore, the ANN-linked LM is a potent chemometrics method because of its high performance and good prediction for non-linear systems (Ghaedi, 2015). The typical network architecture of AI & D/ML or ANN is organized in three-different layers, viz., one input, one output and one or more hidden layers. Figure 2 portrays the schematic architecture of AI & D/ML or ANN having three input, ten hidden and three output layers (3:10:3). The architectural structure of AI & D/ML or ANN is the most common multi-layered perceptron (MLP) type which is built on four different elements, input, hidden and output layers along with connections or weights. Interestingly, the MLP type AI & D/ML or ANN works in two phases, training and testing. The training phase is based on the iterative demonstration of the available data pattern to teach the AI & D/ML or ANN for accomplishing the designated assignment. Figure 3 shows the various steps involved in developing the neural network. Other frequently used neuronal network combinations are the kohonen network, convolutional neural network (CNN) and recurrent neural network (RNN).

[Insert Figs. 2 & 3 here]

Figure 4 depicts the possible way to integrate AI & D/ML or ANN in the optimization of analytical method conditions and formula of the dosage form. It can be seen from Figure 4 that the AI & D/ML or ANN needs to be integrated with the DoE approach for optimizing conditions for analytical method and formula for the dosage form. This type of integration between AI & D/ML or ANN and DoE allows the coining of new terminology called, “double-stage systematic optimization”. The double-stage systematic optimization was therefore started initially by using the conventional DoE approach and then by the application of AI & D/ML or ANN. For instance, Rahman et al., (2021) have used the RSM generated from face-centered CCD of DoE while the ANN is linked with the LM algorithm of AI & D/ML or ANN. Table 3 displays selected non-comprehensive publications showing the involvement of AI & D/ML or ANN in the optimization of analytical method conditions and the formula of the dosage form.

[Insert Figure 4 & Table 3 here]

Because of AI & D/ML or ANN's advantage in dealing with complex and unstructured data, it is well suited for addressing a wide range of applications in the pharmaceutical sciences and easing up the process (Simões et al., 2020). Table 4 displays the various algorithms usage coupled with AI & D/ML or ANN in different pharmaceutical product development processes.

[Insert Table 4 here]
Representing the drug release process by using computationally simple empirical models is a challenging task since there are complicated interactions between formulation and processing variables. The effort in the pre-prescription step would be considerably reduced if AI & D/ML or ANN model could forecast drug release, and the accuracy of the predictions has been proven. Nagy et al., (2019) used the near-infrared (NIR) and Raman spectra to compare four three-layer ANN models to the standard partial least square (PLS) regression to predict the dissolution profile of extended-release anhydrous caffeine tablets. Brahima et al., (2017) used an MLP for the modelling of riboflavin release behaviour from poly(NIPA-co-AAc) hydrogels. The results showed that the function of ANN was validated, and when compared to the RSM using the mean square error (MSE), the ANN was more appropriate for predicting the release of riboflavin hydrogels and had great generality over the release behaviour of hydrogel. Additionally, Elman neural networks (ENNs) and other dynamic neural networks can also be used to forecast dissolution profiles. Petrovic et al., (2012) used both an ENN and an MLP to characterise the release curve of tablets and determined the wide applicability of ANN. Husseini et al., (2009) and Moussa et al., (2017) used ANN to optimise the ultrasonic release of APIs in preparations (such as liposomes and micelles) to keep therapeutic concentrations constant at specific sites. Han et al., (2018) predicted the disintegrating time of disintegrating oral tablets by neural network techniques. A few selected examples of analytical method condition optimization and formula optimization by integrating the AI & D/ML or ANN concept are narrated below.

**SVM in formulation development**

Wang et al. (2022) used particle swarm optimization along with the least square support vector machine (PSO-LSSVM) to simplify the optimization process. The results of the prediction model and Taguchi design were compared with PSO-LSSVM. Additionally, this model provided lower costs and a more efficient design of pharmaceutical formulation.

**SVM in analytical method development**

Keyvan et al., (2021)suggested UV spectrophotometric method development using feed-forward artificial neural network (FFNN) and least square support vector machine (LS-SVM) to simultaneously investigate Sofosbuvir and Daclatasvir in tablet production and biological fluid. Results indicated that the technique has a high potential for predicting component concentrations in dosage forms with a shorter analysis time.

**GA in formulation development**

Kumar and Kumar (2019) used integral hybrid GA with BPANN and RSM based on a central composite design considering water fraction, surfactant fraction, powder density and ultrasonication time as analysing parameters. Results indicated that the multi-objective hybrid GA model was successful in establishing robust results compared to the conventional method.

**GA in analytical method development**
In the study by Attia et al., (2021) GA-ANN was used to quantitatively analyse the UV absorption of velpatasvir and sofosbuvir which revealed some overlap, indicating difficulty in the simultaneous estimation of two drugs. GA-ANN proved to be effective in estimating drugs, with acceptable values of root mean square errors for calibration and prediction.

**Autoencoder in analytical method development**

Kensert et al. (2021) developed a deep one-dimensional convolutional autoencoder that simultaneously eliminates baseline noise and baseline drift to detect and quantify analytes in a mixture of chromatograms with high number and diversity surpassing the approaches like Savitzky-Golay smoothing, Gaussian smoothing and wavelet smoothing.

**AI & D/ML or ANN in optimizing eutectic solvent system**

The use of AI & ML or ANN to predict and select solvent systems is a very interesting integration of academia and industry. AI & D/ML and ANN can assist in optimizing eutectic systems by designing a solvent system based on appropriate properties. By involving AI & D/ML or ANN-based algorithms, the eutectic solvent system may be chosen by automatically separating the products from the reaction solution (self-precipitation) (Amar et al., 2019; Von Lilienfeld, 2018).

**Conclusion**

The DoE (linear model)-supported AI & D/ML or ANN concept is currently being established to optimize not only the analytical method conditions for single or dual drug quantification but also the formula development for dosage form. On the other hand, additional ways are also identified in recent years to omit the linear model support for AI & D/ML or ANN for intriguing of pharmaceutical product development processes. To go ahead directly with non-linear models, the employment of multiple-input-multiple-output and multiple-input-single output-based ANN architecture, self-organizing map, MLP network trained with backpropagation algorithm, three- or multi-layer feed-forward ANN, etc. are being proposed currently by Yun et al., (2020) and Ramakrishna et al., (2014) Such kind of direct involvement of AI & D/ML or ANN will speed up the production performance as well as reduce human errors leading to improve the quality of pharmaceutical products.

**Abbreviations**

APIs- active pharmaceutical ingredients

AAE- adversarial autoencoder

AI & D/ML- artificial intelligence and deep/machine learning

AI-DSS- artificial intelligence-based decision support system
ANN- artificial neural network
AE- autoencoder
BPANN- back propagation artificial neural network
BLIS- bacteriocin-like inhibitory substances
CCD- central composite design
Chk 1- Checkpoint kinase 1
CGB- comprehensive grobner bases
CWT- continuous wavelet transform
CAAs- critical analytical attributes
CS- cuckoo search
DBN- deep belief network
DoE- design of experiment
dbCICA- docking-based comparative intermolecular contacts analysis
DruGAN- drug generative adversarial network
ENNs- elman neural network
FFNN- feedforward neural network
GA- genetic algorithm
IPO- initial public offering
kNN- k nearest neighbour
LS-SVM- least square support vector machine
LM- Levenberg-Marquardt
LC-MS/MS- liquid chromatography-mass spectrometry/mass spectrometry
LSTM- long short-term memory
MSE- mean square error
MLFN-12- multilayer feed forward neural network model

MLP- multilayer perceptron

MLR- multilinear regression

NB- naive bayes

NCE- new chemical entity

NIR- near infrared

OFAT- one-factor-at-a-time

ODT- oral disintegrating tablet

PLS- partial least square

PSO- particle swarm optimization

PBD- Plackett-Burman design

PAT- process analytical technology

QbT- quality by testing

RF- random forest

RNN- recurrent neural network

RSD- relative standard deviation

RSM- response surface methodology

RBM- restricted Boltzmann machine

RP-HPLC- reverse phase-high performance liquid chromatography

SNEDDS- self-nanoemulsifying drug delivery system

SMILES- simplified molecular-input line-entry system

CSTR- single continuous-stirred tank reactor

SNARC- stochastic neural analogue reinforcement computer

SVM- support vector machine
Declarations

Acknowledgements

The funding was provided by Department of Pharmaceuticals, Ministry of Chemicals and Fertilizers, Govt. of India to the authors Naitik Jariwala, Chandra Lekha Putta, Ketki Gatade, Manasi Umarji, Syed Nazrin Ruhina Rahman, Amoolya Sree, Atul Sayaji Kamble & Abhinab Goswami and by National Mission on Himalayan Studies (NMHS), Ministry of Environment, Forest & Climate Change (MoEF & CC) Nodal and Serving hub with G.B. Pant National Institute of Himalayan Environment & Sustainable Development, Govt. of India to Datta Maroti Pawde are greatly acknowledged.

CRediT (Contributor Roles Taxonomy) author statement

Naitik Jariwala: Conceptualization, Methodology, Software; Chandra Lekha Putta: Conceptualization, Methodology, Software; Ketki Gatade: Conceptualization, Methodology, Software; Manasi Umarji: Conceptualization, Methodology, Software; Syed Nazrin Ruhina Rahman: Data curation, Writing- Original draft preparation; Datta Maroti Pawde: Project administration; Amoolya Sree: Visualization, Investigation; Atul Sayaji Kamble: Visualization, Investigation; Abhinab Goswami: Software, Validation; Tamilvanan Shunmugaperumal: Writing- Reviewing and Editing, Supervision.

Conflict of Interest

The authors declare that they have no competing interest.

Funding Statement

There is no funding source.

References


## Tables

**Table 1** Historical overview on artificial intelligence & deep/machine learning or artificial neural network in ascending order of milestone years [Partially taken from Manikiran & Prasanthi, 2019]

<table>
<thead>
<tr>
<th>Milestone Years</th>
<th>Landmark of artificial intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1943</td>
<td>Walter Pitts and Warren Mcculloch proved that the neurons can do logical operations like “and”, “or” or “not” when they are connected as a network.</td>
</tr>
<tr>
<td>1951</td>
<td>Marvin Minsky introduced the first neuronal network i.e., SNARC (Stochastic Neural Analogue Reinforcement Computer) which can solve the problem from the real world.</td>
</tr>
<tr>
<td>1956</td>
<td>The term artificial intelligence was coined at Dartmouth College Conference.</td>
</tr>
<tr>
<td>1958</td>
<td>Frank Rosenblatt created perceptrons which are neuronal networking that transmit information in one direction which is the basis of today’s AI progress.</td>
</tr>
<tr>
<td>1969</td>
<td>The symbolic representation of the problems is supported by Minsky in his book “perceptrons”.</td>
</tr>
<tr>
<td>1974-1980</td>
<td>“First AI winter” period.</td>
</tr>
<tr>
<td>1986</td>
<td>Georey Hinton developed back propagation algorithm design which is used in deep learning these days.</td>
</tr>
<tr>
<td>1997</td>
<td>IBM deep blue defeated the Garry Kasparov who is a Russian grandmaster in chess.</td>
</tr>
<tr>
<td>2013</td>
<td>British technology is used by the Google to perform efficient research on photos.</td>
</tr>
<tr>
<td>2016</td>
<td>Google DeepMind defeated the Go Champion Lee Sedol by the software alpha Go.</td>
</tr>
<tr>
<td>2017</td>
<td>Use of AI in analytical method optimization of pharma products.</td>
</tr>
<tr>
<td>2018</td>
<td>AI-based formula optimization for pharmaceutical products.</td>
</tr>
<tr>
<td>2019</td>
<td>Integration DoE with AI.</td>
</tr>
<tr>
<td>2020</td>
<td>AI concept in pharmaceutical unit operations including process analytical technology (PAT).</td>
</tr>
<tr>
<td>2021</td>
<td>AI utility in pharmaceutical labeling and package security system.</td>
</tr>
</tbody>
</table>

**Table 2** Non-comprehensive financial dealings of pharma companies related to the application of artificial intelligence & deep/machine learning or artificial neural network in drug discovery process [Partially taken from Savage, (2021)]
<table>
<thead>
<tr>
<th>Companies</th>
<th>Month &amp; Year</th>
<th>Headline</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZebiAI Therapeutics, Relay Therapeutics</td>
<td>April 2021</td>
<td>Relay Therapeutics buys ZebiAI Therapeutics for $85 million and further $185 million in potential revolutionary payments.</td>
</tr>
<tr>
<td>Recursion Pharmaceuticals (Salt Lake City, UT, US)</td>
<td>April 2021</td>
<td>Recursion Pharmaceuticals finalizes $436 million initial public offering (IPO).</td>
</tr>
<tr>
<td>Iktos (Paris, France), Pfizer (New York City, USA)</td>
<td>March 2021</td>
<td>A number of Pfizer's small-molecule programs are supported by Iktos by its AI-driven <em>de-novo</em> design software.</td>
</tr>
<tr>
<td>Exscientia (Oxford GB)</td>
<td>March 2021</td>
<td>Exscientia finalizes $100 million Series C financing, with investors including Evotec, Bristol Myers Squibb and GT Healthcare.</td>
</tr>
<tr>
<td>Insitro (South San Francisco, CA, US)</td>
<td>March 2021</td>
<td>Canada Pension Plan Investment Board supported in financing Insitro to raise $400 million in Series C.</td>
</tr>
<tr>
<td>Valo Health (Boston, MA, USA)</td>
<td>March 2021</td>
<td>Valo Health, which is emerging its Opal computational drug discovery and development platform, increases $110 million to add to its $190 million upturned in January 2021 for its Series B funding round.</td>
</tr>
<tr>
<td>Roivant (Basel, New York), Silicon Therapeutics (Boston, USA)</td>
<td>February 2021</td>
<td>Silicon Therapeutics was sold to Roivant for $450 million, including its physics-based platform for insilico small-molecule drug design, to be combined with Roivant's machine learning approaches.</td>
</tr>
<tr>
<td>Cellarity (Cambridge, MA)</td>
<td>February 2021</td>
<td>For the drug discovery approach related to modulating cellular behaviors, Cellarity raised $123 million in Series B funding.</td>
</tr>
<tr>
<td>AbCellera (Vancouver, Canada)</td>
<td>December 2020</td>
<td>AbCellera ends its IPO at $556 million.</td>
</tr>
<tr>
<td>Insitro (South San Francisco, CA, US), Bristol Myers Squibb (New York, USA)</td>
<td>October 2020</td>
<td>Insitro Human platform, a machine learning technology is used to detect potential drug targets by rising prognostic models of amyotrophic lateral sclerosis and frontotemporal dementia. Then, Bristol Myers Squibb will select candidates to progress further.</td>
</tr>
<tr>
<td>Company</td>
<td>Date</td>
<td>Event Description</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Recursion Pharmaceuticals (Salt Lake City, UT, US), Bayer (Leverkusen, Germany)</td>
<td>September 2020</td>
<td>Deal between Bayer partners and Recursion Pharmaceuticals in drug development used Recursion's AI-guided drug discovery platform for new small-molecule therapies for treating fibrotic diseases causing addition in Series D financing round.</td>
</tr>
<tr>
<td>Recursion Pharmaceuticals (Salt Lake City, UT, US)</td>
<td>September 2020</td>
<td>Recursion Pharmaceuticals, apply machine learning to cellular imaging data and raises $239 million in Series D financing round directed by Bayer's investment department Leaps. Other investors include Baillie Gifford, Casdin Capital, Samsara BioCapital and Lux Capital.</td>
</tr>
<tr>
<td>XtalPi (Cambridge, Massachusetts, US)</td>
<td>September 2020</td>
<td>XtalPi is using quantum physics with AI to discover drug candidates and more than a dozen investment companies raised $318 million in Series C round for this start-up.</td>
</tr>
<tr>
<td>BioSymetrics (New York City, USA), Sema4 (Stamford, Connecticut, US), Johnson &amp; Johnson (New Jersey, United States)</td>
<td>August 2020</td>
<td>J&amp;J's Janssen is collaborated with BioSymetrics and Sema4 aiming to apply machine learning to predict onset and severity of COVID-19 among different populations, involved developing new treatments and vaccines.</td>
</tr>
<tr>
<td>Atomwise (San Francisco, CA, US)</td>
<td>August 2020</td>
<td>Sanabil Invests around $123 million Series B funding for Atomwise in order to upkeep the advancement of its molecule identification software.</td>
</tr>
<tr>
<td>Relay Therapeutics (Cambridge, USA)</td>
<td>July 2020</td>
<td>Relay Therapeutics, focused on protein motion to design drug candidates, calls $400 million initial public offering (IPO).</td>
</tr>
<tr>
<td>Insitro (South San Francisco, CA, US)</td>
<td>May 2020</td>
<td>Insitro for its machine learning-based drug discovery approach, raised $143 million in Series B funding.</td>
</tr>
<tr>
<td>AbCellera (Vancouver, Canada)</td>
<td>May 2020</td>
<td>AbCellera for the expansion of its antibody drug discovery platform raised $105 million in Series B funding.</td>
</tr>
<tr>
<td>Schrödinger (New York City, USA)</td>
<td>February 2020</td>
<td>Bill Gates and David Shaw's Drug discovery software company ends $232 million initial public offering (IPO).</td>
</tr>
<tr>
<td>Exscientia (Oxford GB), Bayer (Leverkusen, Germany)</td>
<td>January 2020</td>
<td>Bayer partnered with Exscientia to discover cardiovascular and oncology drug candidates by their artificial intelligence drug discovery platform i.e. “Centaur Chemist”.</td>
</tr>
<tr>
<td>ZebiAl Therapeutics, Google Accelerated Science</td>
<td>January 2020</td>
<td>ZebiAl Therapeutics partnered with Google Accelerated Science for discovering small-molecule drug candidates using machine learning and DNA-encoded library technologies.</td>
</tr>
<tr>
<td>Microsoft, Novartis (Basel Switzerland)</td>
<td>October 2019</td>
<td>Novartis announces collaboration with Microsoft to apply their AI algorithms to its large datasets in order to identify and develop therapeutics.</td>
</tr>
<tr>
<td>Exscientia (Oxford GB), Rallybio (New Haven, CT)</td>
<td>July 2019</td>
<td>Exscientia collaborates with Rallybio bringing its AI drug discovery platform Centaur Chemist to discover small molecule drugs for rare diseases.</td>
</tr>
</tbody>
</table>
Insitro (South San Francisco, CA, US), Gilead (Foster City, California, USA)  
April 2019  
Gilead is the first big pharma company to sign 3-year deal with Insitro for applying its Insitro Human platform to identify new drug targets for non-alcoholic steatohepatitis by producing experimental models of the disease. Insitro will be paid $15 million in a deal potentially worth $1 billion.

BenevolentAI (London United Kingdom), AstraZeneca (Cambridge, United Kingdom)  
April 2019  
AstraZeneca collaborated with BenevolentAI to detect new drug candidates for chronic kidney disease and idiopathic pulmonary fibrosis.

Table 3 Selected non-comprehensive publications showing the involvement of AI & D/ML or ANN in optimization of analytical method conditions and formula of dosage form
<table>
<thead>
<tr>
<th>Publication title</th>
<th>Short summary of the publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application of design of experiments approach-driven artificial intelligence and machine learning for systematic optimization of reverse phase high performance liquid chromatography method to analyze simultaneously two drugs (cyclosporin A and etodolac) in solution, human plasma, nanocapsules and emulsions (Rahman et al., 2021).</td>
<td>- Optimization of the RP-HPLC method conditions using CCD-driven ANN linked with Levenberg-Marquardt (LM) models (Rahman et al., 2021).</td>
</tr>
<tr>
<td>The application of artificial neural network and least square support vector machine methods based on spectrophotometry method for the rapid simultaneous estimation of triamcinolone, neomycin and nystatin in skin ointment formulation (Abasi et al., 2021).</td>
<td>- ANN and LS-SVM were employed (Abasi et al., 2021).</td>
</tr>
<tr>
<td>Artificial neural networks in tandem with molecular descriptors as predictive tools for continuous liposome manufacturing (Sansare et al., 2021).</td>
<td>- The use of molecular descriptors as inputs in ANN has the potential to reduce predictive error and thus improve prediction accuracy (Sansare et al., 2021).</td>
</tr>
<tr>
<td>Cocrystal prediction using machine learning models and descriptors (Mswahili et al., 2021).</td>
<td>- SMILES descriptor values were developed and machine learning models were compared. ANN has the best accuracy, sensitivity and F1 score (Mswahili et al., 2021).</td>
</tr>
<tr>
<td>Retention time prediction using neural networks increases identifications in crosslinking mass spectrometry (Giese et al., 2021).</td>
<td>- The xiRT Siamese machine learning model predicts retention time with great accuracy (Giese et al., 2021).</td>
</tr>
<tr>
<td>Predicting drug release from diazepam fused deposition modeling printed tablets using deep learning approach: Influence of process parameters and tablet surface/volume ratio (Obeid et al., 2021).</td>
<td>- An established ANN model could predict drug release features (Obeid et al., 2021).</td>
</tr>
<tr>
<td>Development of LC-MS/MS determination method and backpropagation artificial neural networks pharmacokinetic model of febuxostat in healthy subjects (Xu et al. 2021).</td>
<td>- Back propagation artificial neural network (BPANN) model for forecasting febuxostat plasma concentrations was constructed (Xu et al. 2021).</td>
</tr>
<tr>
<td>Application of the artificial neural network to optimize the formulation of self-nanoemulsifying drug delivery system (SNEDDS) containing rosvastatin (Vu et al., 2020).</td>
<td>- The authors were able to improve the formulation of rosvastatin SNEDDS using an artificial neural network (Vu et al., 2020).</td>
</tr>
<tr>
<td>Modelling the absorbance of a bioactive compound in HPLC method using artificial neural network and multilinear regression (MLR) methods (Abdullahi et al., 2020).</td>
<td>- The ANN model performed marginally better</td>
</tr>
</tbody>
</table>
Simultaneous spectrophotometric estimation of atenolol and amlodipine besylate in pharmaceutical formulations (Hasan et al., 2020).

- ANN, PLS and principal component regression are three multivariate analytic approaches that have been created (Hasan et al., 2020).

**Table 4** Various algorithms usage in different pharmaceutical product development processes
<table>
<thead>
<tr>
<th>Algorithm name</th>
<th>Algorithm functioning</th>
<th>Application in pharma industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support Vector Machine (SVM)</td>
<td>According to Patel et al., (2020), the SVMs are managed machine learning algorithms which differentiate compounds into groups based on certain attribute selection by originating a hyperplane. SVM is a statistical method for mapping data into high-dimensional space to identify a lower dimensional hyperplane that maximizes the data separation using a nonlinear kernel (Lo et al., 2018).</td>
<td>Stephenson et al., (2019) utilized the SVM algorithm along with other algorithms to classify resistivity and susceptibility of tuberculosis drugs.</td>
</tr>
<tr>
<td>Autoencoder (AE)</td>
<td>AE is utilized layer by layer as an elementary unit for developing deep neural network (Ghasemi et al., 2018).</td>
<td>Kadurin et al., (2017) implemented DruGAN (Drug Generative Adversarial Network) which is set up on AAE (Adversarial Autoencoder) model to recognize novel molecular compounds having anticancer activity.</td>
</tr>
<tr>
<td>k Nearest Neighbor (kNN)</td>
<td>Ferreira &amp; Andricopulo, (2019) outlined that kNN as a non-linear correlation optimization method which creates its own understanding from acquired knowledge of data.</td>
<td>Jaradat et al., (2015) applied combination of kNN with dbCICA (Docking-based comparative intermolecular contacts analysis) tool for selection and identification of anticancer drug as associated with Chk 1 (Checkpoint Kinase 1) inhibitors.</td>
</tr>
<tr>
<td>Random Forest (RF)</td>
<td>According to Patel et al., (2020), RF is the most suitable for larger database having several attributes. RF is a collective learning method of various decision tree assembled on basis of different datasets and by using commonly elected program (Lo et al., 2018).</td>
<td>Lee et al., (2017) applied RF based virtual screening method to develop QSAR models with support of ChEMBL datasets.</td>
</tr>
<tr>
<td>Genetic algorithm (GA)</td>
<td>GA is an improvised algorithm developed to discover ideal result based on procedure with different steps (Ghodmare et al., 2018).</td>
<td>AutoGrow4: It is based on genetic algorithm and aids in prediction of drug-like compounds and optimizing pre-existing ligands. The process involves seed selection and pooling, then crossing over and filtering of compound and lastly it docks the compound and ranks them based on scoring (Spiegel &amp; Durrant, 2020).</td>
</tr>
<tr>
<td>Deep belief network (DBN)</td>
<td>DBN is classified as generative unsupervised learning technique, developed by assembling restricted boltzmann machine</td>
<td>DBN algorithm is able to establish a classification system into drug like and non-drug like potential druggable candidates from ZINC datasets to</td>
</tr>
</tbody>
</table>

Page 24/29
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Description</th>
<th>Related Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBM</td>
<td>(RBM) and is instructed in an acquisitive design (Hooshmand et al., 2021).</td>
<td>Accelerate drug discovery process (Hooshmand et al., 2021).</td>
</tr>
<tr>
<td>Naive Bayes (NB)</td>
<td>A probabilistic classification method that practices probability prior and Bayes’ theorem to predict membership by assuming feature independency (Lo et al., 2018).</td>
<td>Zhang et al., (2017) utilized NB classifier model for determining the mitochondrial toxicity caused by drugs.</td>
</tr>
<tr>
<td>Levenberg-Marquardt (LM)</td>
<td>LM algorithm is a very popular nonlinear optimization method (Ramadasan et al., 2017).</td>
<td>Valizadeh et al., (2021) performed simultaneous estimation of anti-asthmatic drugs in inhalation spray using LM.</td>
</tr>
<tr>
<td>Partial least square (PLS) algorithm</td>
<td>According to Wang et al., (2019), PLS is an efficient multivariate statistical data analysis method that has been proven to obtain good prediction accuracy which has gain numerous applications in bioinformatics, machine learning and chemometrics.</td>
<td>Tinkov et al., (2021) executed different types of molecular descriptors and machine learning methods by constructing QSAR models using PLS.</td>
</tr>
<tr>
<td>Backpropagation artificial neural network (BPANN)</td>
<td>Sherwani et al., (2021) applied the BPANN aimed to expedite the convergence of neural network to global optimum</td>
<td>Ajdarić et al., (2021) predicted stability of esomeprazole 40 mg freeze dried powder for solution using deep learning approach. They trained multilayer perceptron (MLP) network using back propagation network.</td>
</tr>
<tr>
<td>Cuckoo search algorithm</td>
<td>According to Mareli and Twala (2018), Cuckoo search is nature-inspired algorithm used extensively to solve optimization problems because it can maintain balance between local and global random walks using switching parameter.</td>
<td>Real-time detection of covid-19 using cuckoo search optimized extreme learning machine was done by Hu et al., (2021) using chimp optimization algorithm.</td>
</tr>
<tr>
<td>Recurrent Neural Network (RNN) algorithm</td>
<td>RNNs is a sub-class of ANNs, and they are structured to use hidden variables as a memory for capturing temporal dependencies between system and control variables (Wong et al., 2018).</td>
<td>Wong et al., (2018) demonstrated the use of RNNs for system identification of a complex reaction in a single continuous-stirred tank reactor for pharmaceutical API production.</td>
</tr>
</tbody>
</table>

**Figures**
Figure 1

Applications of AI & D/ML or ANN in pharmacy field of healthcare system
Figure 2

Schematic architecture of ANN having three input, ten hidden and three output layers
Figure 3

Schematic representation of development of an AI & D/ML or ANN
Figure 4

Integration of AI & D/ML or ANN in analytical method development and formula optimization process