REVIEW ARTICLE

Theme: Process Analytical Technology (PAT) in Pharmaceutical and Biopharmaceutical Industry

Application of Artificial Neural Networks in the Process Analytical Technology of Pharmaceutical Manufacturing—a Review

Brigitta Nagy[1] · Dorián László Galata[1] · Attila Farkas[1] · Zsombor Kristóf Nagy[1]

Abstract

Industry 4.0 has started to transform the manufacturing industries by embracing digitalization, automation, and big data, aiming for interconnected systems, autonomous decisions, and smart factories. Machine learning techniques, such as artificial neural networks (ANN), have emerged as potent tools to address the related computational tasks. These advancements have also reached the pharmaceutical industry, where the Process Analytical Technology (PAT) initiative has already paved the way for the real-time analysis of the processes and the science- and risk-based flexible production. This paper aims to assess the potential of ANNs within the PAT concept to aid the modernization of pharmaceutical manufacturing. The current state of ANNs is systematically reviewed for the most common manufacturing steps of solid pharmaceutical products, and possible research gaps and future directions are identified. In this way, this review could aid the further development of machine learning techniques for pharmaceutical production and eventually contribute to the implementation of intelligent manufacturing lines with automated quality assurance.

KEY WORDS artificial neural network · machine learning · Pharma 4.0 · Process Analytical Technology · real-time release testing

INTRODUCTION

In the new century, manufacturing industries have started to undergo a vast transformation, fueled by digitalization, automation, and the tremendous amount of data collected during the manufacturing processes. Also referred to as the Fourth Industrial Revolution (Industry 4.0 (1) or Pharma 4.0 in the pharmaceutical industry (2, 3)), the evolving digital technology includes data-driven manufacturing and the vision of smart factories, where interconnected systems can communicate with each other and make autonomous decisions (4).

In the pharmaceutical industry, modernization is also promoted by the Quality by Design (QbD) (5) and the Process Analytical Technology (PAT) (6) frameworks. The QbD approach emphasizes the need for product and process understanding, *i.e.*, the identification of the critical material attributes (CMAs) and critical process parameters (CPPs) that significantly influence the critical quality attributes (CQAs) of the product and process. This leads to establishing a design space, within which the quality is deemed acceptable. Consequently, the flexibility of production can be increased, as operating within a regulatory approved design space is not regarded as a change. The PAT initiative also aims the science- and risk-based production by emphasizing the need for the real-time measurements of the CQAs and CPPs with in-process sensors, coupled with the corresponding data analysis methods and control strategy. As the QbD and PAT principles are being implemented, the possibility of reducing the labor-intensive and time-consuming quality control test on the final products is also becoming a reality. Instead, the real-time release testing (RTRT) approach can be used, that is, the process and product understanding and

✉ Zsombor Kristóf Nagy
zsknagy@oct.bme.hu

1 Department of Organic Chemistry and Technology, Faculty of Chemical Technology and Biotechnology, Budapest University of Technology and Economics, Műegyetem rkp. 3., Budapest H-1111, Hungary

the adequate real-time monitoring and control of the process can serve as the quality assurance (7). Consequently, the QbD, PAT, and RTRT concepts—along with the advancements in data processing and automation—are indispensable for the aimed agile and innovative manufacturing, and their implementation could eventually lead to smart and eventually autonomous pharmaceutical factories (3).

To realize the aims of Industry/Pharma 4.0, artificial intelligence (AI) and machine learning (ML) techniques have emerged as versatile tools (2, 8) to tackle several arising tasks, *e.g.*, the analysis of big data (9) or the development of digital twins (the digital counterpart of a physical system) (10, 11). Artificial intelligence is mainly referred to as the computational methods that perform tasks typically associated with human-like thought processes, such as pattern recognition or decision-making. Within AI, machine learning techniques accomplish these tasks by learning from a provided dataset to produce a response without being explicitly programmed to do so (12). ML can adjust the model behavior to continuously improve its performance as the training dataset is expanding, which makes it especially suitable for data-driven manufacturing purposes. The medical regulatory agencies also show increasing openness for AI/ML approaches. For example, the Danish Medicines Agency has recently published a list of questions to consider for developing and applying ML-based models in GxP-regulated areas (13). Furthermore, the US Food and Drug Administration has issued an action plan for establishing a "Good Machine Learning Practice" (14). Although it deals with ML-based medical devices, the approach (*e.g.*, how to evaluate robustness, bias, and real-world performance) could be potentially further generalized for other ML applications.

ML comprises several different mathematical approaches, such as artificial neural networks (ANNs), deep learning, support vector machines (SVM), and decision trees (12, 15). This work primarily focuses on ANNs, which have gained tremendous attention due to their flexibility in describing complex, linear, or non-linear relationships for different purposes, such as pattern recognition, regression, or time-series forecasting. ML techniques have already found several applications in the different stages of pharmaceutical research and development, such as for target selection, clinical trials, quantitative structure-activity relationships studies, or formulation optimization (16–21).

However, in preparation for smart manufacturing, the real-time applicability of ML in the pharmaceutical manufacturing processes also needs to be studied, which, to the best of the authors' knowledge, has not been studied in detail in previous review papers yet. Therefore, this paper aims to explore the current state of ML techniques (mostly ANNs) within the PAT framework. The application of ML together with analytical sensor systems for process monitoring and control purposes is reviewed, considering the most common

upstream and downstream manufacturing steps of small molecule active pharmaceutical ingredients (APIs) and solid pharmaceutical formulations (*e.g.*, tablets or capsules). This overview also aims to identify potential research directions, future challenges, and risks associated with implementing ANNs within PAT. Consequently, this review could facilitate the development of smart pharmaceutical manufacturing approaches and aid the digitalization efforts of the pharmaceutical industry.

ARTIFICIAL NEURAL NETWORKS

The development of ANNs was inspired by the information processing behavior of the human brain, as the calculation is based on interconnected information processing units, *i.e.*, artificial neurons (also called nodes or perceptrons), which receive inputs and convert them to desired outputs (Fig. 1a). This is achieved by first weighting and summarizing the inputs by an activation function and then calculating the outputs using a predefined transfer function (see Fig. 1b). In such a way, the information is passed through numerous neurons to produce the final output (Fig. 1c).

Depending on the purpose of the model, arbitrary NN topologies can be built, *e.g.*, by varying the number of neurons, their connections to each other, and the applied transfer functions. In most applications, the nodes with the same tasks are organized into layers. A standard topology is the feedforward neural network (Fig. 1c)—or multilayer perceptron (MLP)—where the information passes through the network without any back loops to previous layers. The nodes in the input layer receive the information from the outside world and pass it to further nodes. The actual calculations happen in the hidden layer(s), which have no direct connection to the outside world but pass the information to the output nodes. The output layer, where the final transformations happen, provides the network results. Loops are also possible within the network, creating feedback, or recurrent neural networks (RNN), where the information also travels back to previous processing units, which exhibits a memory-like behavior. This is especially important for analyzing, *e.g.*, sequential and time-series data. Several RNNs have been developed for different purposes, such as the Elman network, layer recurrent neural network (LRNN), non-linear autoregressive exogenous model (NARX), the long short-term memory (LSTM) NN (22).

According to the universal approximation theorem (23), a network with one hidden layer with a finite number of neurons can estimate any continuous function with arbitrary transfer functions. However, the utilization of multiple hidden layers (called a deep neural network if it contains at least three hidden layers) can also be necessary or more effective for complex tasks, *e.g.*, for processing unstructured
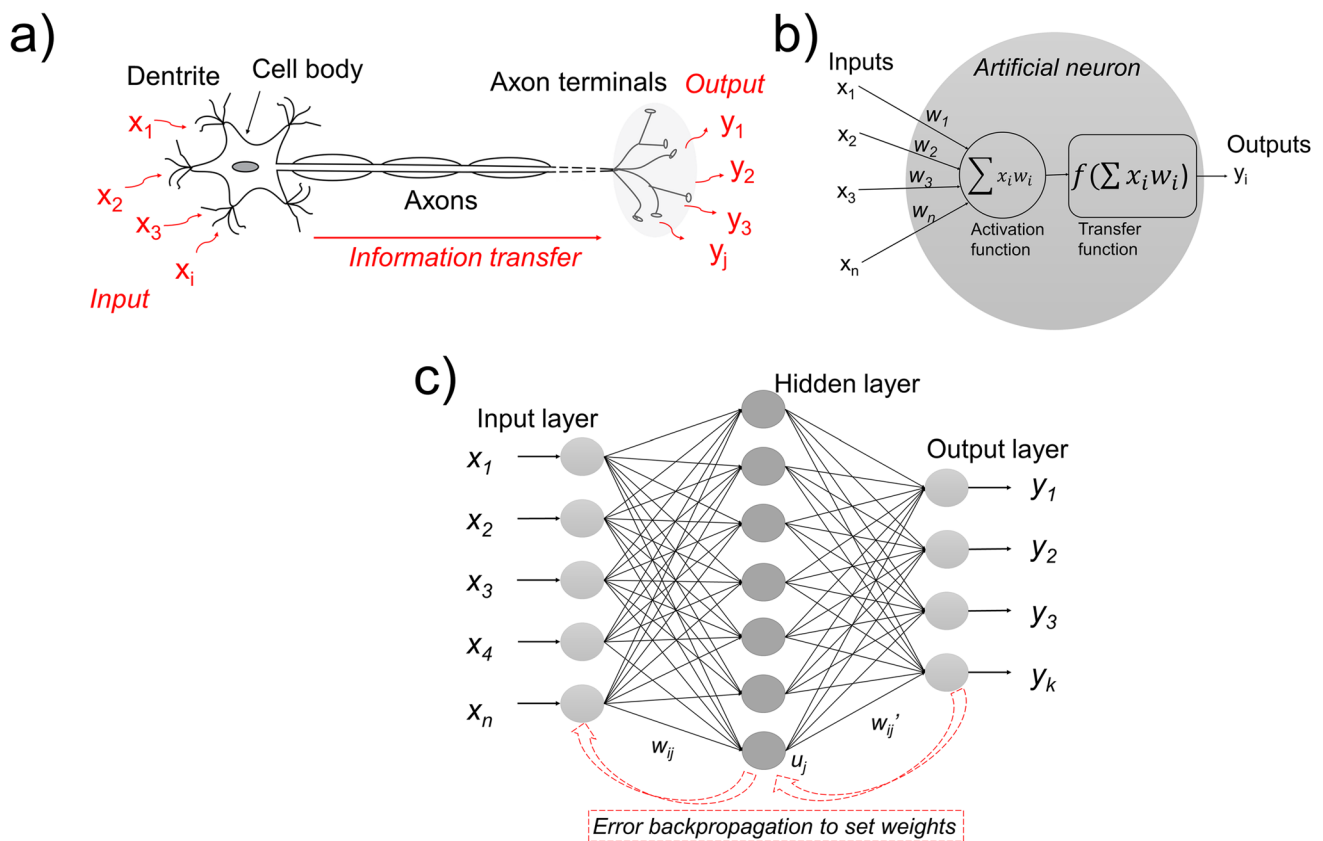
Fig. 1 Representation of **a** human neuron, **b** a single artificial neuron, and **c** a multi-layered feedforward backpropagation (FF-BP) neural network

data. Most famous deep networks include, *e.g.*, RNNs (22) and convolutional neural networks (CNN) (24). CNNs are mainly used in image analysis for segmentation, classification, and object detection. They contain several hidden layers for different purposes, such as convolutional layers for feature extraction, pooling layers for reducing the dimension, and fully connected layers for classification or making a prediction. As training CNNs from scratch can have vast computational and data demand, several pre-trained CNN architectures are available, such as the AlexNet, GoogLeNet, or ResNet model, consisting of 8, 22, and 152 layers, respectively. These models can be utilized for transfer learning, *i.e.*, using them as the starting point (*e.g.*, for feature extraction) to model a new problem and to tune the model to the required training data. For more information on deep learning, CNNs, and transfer learning, see (24, 25).

Besides the network topology, the nodes' transfer functions significantly determine the model's behavior. Some of the frequently used transfer functions are the linear (often used in the output neurons for function fitting), the log-sigmoid (provides output between 0 and 1, used for discrete and binary outputs, *e.g.*, in pattern recognition problems), and tangent sigmoid function (resulting in an output from

−1 to +1, often used for regression tasks). Another widely used transfer function for deep NNs is the rectified linear unit (ReLU), which does not change the input if it is positive while outputting zero otherwise. Radial basis function (RBF) might also be used, creating RBF networks, which have the advantages of good generalization and the ability to learn in real-time.

ANNs need to be trained for a given task, which means adjusting the weights and biases of the neurons. In supervised learning, this is done using a training dataset consisting of known input-output pairs. Backpropagation is a widely used iterative training approach. First, the weight and biases of the neurons are initialized, either randomly or by following an initialization technique (26) to speed up the learning. Next, an error (cost function), *e.g.*, the mean absolute error (MAE), mean square error (MSE), and the sum of squared error (SSE), is calculated between the network's output and the known target. Based on the obtained error, the training algorithm adjusts the weights and biases, controlled by either a fixed or adaptive learning rate: the lower the rate, a smaller corrective step is taken, which causes longer training but potentially more accurate result. Finally, the cost function is calculated again. One such iteration is

called an epoch, and the process is iterated until a stopping criterion is reached.

The used training algorithms can significantly affect the training time and the performance of the ANN. For example, the gradient descent training algorithm (27) is slow, aiming to find the steepest descent of the error by calculating the first derivatives of the cost function. A widely used method is the Levenberg-Marquardt algorithm (28), the fastest for medium-sized networks but only applicable when the cost function is in the form of the sum of squares. Furthermore, it is sensitive to the weight initialization and prone to overfitting. Bayesian regularization (29, 30) can tackle these problems by expanding the cost function to minimize the linear combination of the squared errors and the sum of squared weights. Consequently, the effective number of the parameters used in the model for the given problem can also be optimized, leading to better generalization and robustness.

APPLICATION OF ANNs IN UPSTREAM PROCESSES

Synthesis

The synthesis of organic compounds is the first step in industrial pharmaceutical manufacturing to produce the API. The process monitoring and control can be utilized in continuous and batch processes to maintain a steady-state, determine endpoint, or optimize operating conditions. ANNs have found applications in the optimization process parameters to improve the outcome and efficacy of the reactions and describe non-linear relationships between spectroscopic data and the desired parameters through black-box multivariate modeling.

The effect of the process parameters (*i.e.*, time, temperature, enzyme amount, molar ratio) on the yield of an enzymatic synthesis of betulinic acid ester could be described by a feedforward (FF) ANN, using 21 training experiments (31). As learning algorithms, quick propagation, incremental backpropagation (BP), batch BP, and Levenberg-Marquardt algorithm were compared, of which the quick propagation gave the best robustness.

Several studies have dealt with the optimization of synthesis using ANNs. For example, Valizadeh *et al.* applied a MLP (32) to optimize the preparation of glucosamine from chitin based on three inputs, *i.e.*, the acid concentration, the acid solution to solid ratio, and the reaction time. The built network was compared to the results of genetic algorithm (GA) and particle swarm optimization methods, which were better in model fitting than the MLP model, but the ANN outperformed them during validation. Optimization of four two-component reactions was also performed by deep reinforcement learning, using RNN (33). The method iteratively found the optimal flow rate, voltage, and pressure of the microdroplet reactions, using fewer steps than other black-box optimization algorithms. ANN-based optimization could also be performed with the combination of computational fluid dynamics (CFD) model, where the CFD model was the source of the training data (34), where several parameters such as conversion, selectivity, and yield were maximized in butadiene synthesis.

RNN could replace a true plant model (35) or state-space model (36) in control algorithms. This could be beneficial to predict the process dynamics within a model predictive control (MPC) of a continuous pharmaceutical synthesis, as the computational demand is significantly decreased compared to the mechanistic models. Moreover, the ANN can greatly benefit from the data-rich environment of the PAT-supported manufacturing.

Although the above examples show that the API synthesis could greatly benefit from ANN modeling, the results are mainly based on historical instead of PAT data. There are a few examples in the literature where the evaluation of inline or on-line PAT measurement by ANNs was presented. For example, in a fermentation, the glucose and the glucuronic acid concentration were determined from Fourier transform infrared (FT-IR) spectroscopic measurements (37) using a multilayer feedforward network with 15 calibration samples. The ANN outperformed the classical partial least squares (PLS) regression. Phenol and chlorophenols were also simultaneously quantified employing an ANN model from UV-Vis spectra, collected by an immersion probe (38). In this case, principal component scores were used to compress the spectra for the training of networks.

Crystallization

Crystallization is crucial in connecting the API synthesis and the downstream formulation steps by providing solid crystalline API, which greatly impacts the final product's yield, purity, further manufacturability, and even bioavailability. PAT sensors are often used, such as ATR-IR, and UV probes can be used to monitor the solute concentration, and focused beam reflectance (FBRM) or *in situ* microscopic measurements (*e.g*., particle vision and measurement, PVM) can indicate the crystal size and count (39).

ML can estimate the crystallization outcome based on historical process data. For example, Velásco-Mejía *et al.* developed ANN and GA models based on the records of 54 industrial batch crystallizations (40). They used nine descriptors and modeled the crystal density as the outcome, which resulted in identifying the most critical parameters and, after optimization, a substantial improvement in the product. In another work, the design space of a cocrystallization process could be explored based on 25 experimental runs and four input variables (41). Using the operating

variables (such as temperature, supersaturation, agitation speed, seeding properties) as ANN inputs, a more accurate crystal growth rate could also be predicted than with multiple non-linear regression (42).

ANNs have been used for extracting information from data-rich PAT tools, such as in-line microscopic images. A ResNet CNN has proved to be effective in classifying crystals detected in PVM images, which was used for contamination classification with >98 % accuracy (43). Such in-line technique can contribute to identifying traces of undesired polymorphs and, therefore, can be used in a feedback control to improve the product purity. Furthermore, the growth rate could also be predicted by measuring the particle size distribution using CNN-based in-line image analysis (44). FBRM measurements provide chord length distribution as particle size information, which, together with the solid concentration, could be used as input for a layer RNN (45) to calculate the crystal size distribution (CSD). Szilágyi and Nagy (46) demonstrated the opposite approach: a direct and fast transformation of two-dimensional CSD (needle-shaped crystals) to chord length distribution and aspect ratio distribution was achieved by a neural network. This was necessary for enabling FBRM and PVM to be used as quantitative direct feedback control tools in a population balance model (PBM)-based control, as the outcome of the PBMs and the analytical sensors are not comparable. The presented approach resulted in 6 times faster calculation than a direct conversion, which could be essential in real-time applications.

ANNs are also getting increasing attention in the control of crystallization. Possible approaches include the self-tuning of the PID controllers by ANN (*e.g.*, by a diagonal RNN (47)) for temperature and level control or the determination of the optimal temperature profile to control the crystal size, *e.g.*, to reduce fines. For the latter, Paengjuntuek *et al.* (48) generated data with a PBM for NN training and then predicted the solution concentration and crystal volume from the temperature and solution concentration data at the previous time points. The ANN was used as a state predictor in the optimization and provided better control performance than conventional methodologies. Furthermore, the trained ANNs have a much lower computational cost than a first-principles model; therefore, they have a great potential to be used in MPC as the predictive model. This has been demonstrated by simulation (49) and experimental studies, using different network types (*e.g.*, feedforward, recurrent, RBF networks) and batch and fed-batch crystallizations (50, 51). However, Öner *et al.* (52) highlighted that mostly historical data were used for model development. In their study, a fully automated laboratory crystallization system has been developed, with temperature and FBRM sensors and using a RBF network. The training was accomplished real-time, using a reference batch and in-line collected data and an updated or a growing data strategy. That is, the network was updated as new experimental data was available. Despite the limited data, the control strategy was robust to various disturbances, such as solvent impurity, seed size, or impeller speed. This approach is applicable even when limited historical data or process understanding is available.

# APPLICATION OF ANNs IN DOWNSTREAM PROCESSES

Powder Blending

The proper execution of powder blending primarily ensures the homogeneous distribution of components in manufacturing solid dosage forms. ML techniques have been applied a few times to assist the real-time analysis of API concentration during a blending process and predict the powders' behavior in various scenarios.

Since the 2000s, it has been shown that predicting the API concentration of powders by ANNs based on near-infrared (NIR) spectra (53, 54) is as effective as the PLS regression. Besides, ANNs can predict the required time to achieve a homogeneous mixture. For example, Tewari *et al.* (55) utilized NIR spectroscopy, ANN, and other multivariate data analysis methods for at-line blending endpoint detection techniques.

Mujumdar *et al.* (56) created a discrete element method model of a sectorial container subjected to oscillations and then simulated the mixing of two particle fractions with different particle sizes under various operating parameters. The simulated results were used as a training dataset to create a FF-BP ANN model—which has much less computational demand—to predict the mean mixing concentration, a parameter describing the effectiveness of the mixing process based on the amplitude of oscillations, the frequency of oscillations, the particle size of the smaller and larger fraction, and the number of cycles. It was concluded that the ANN is an excellent choice when several operating parameters have a non-linear relationship. Furthermore, such techniques could be helpful in the future for control purposes. ANNs could also be applied to process data where the effects of certain factors appear after a time delay. For example, the composition of the blend that leaves the continuous blender could be predicted by an RNN (Fig. 2)—serving as the digital twin of the blender—based on the mass flow rate of the input material streams and the residence time distribution of the system (57). It was found that a non-linear autoregressive network with exogenous inputs can yield results comparable to that of a residence time distribution model.

Fig. 2 Concentration prediction from time-series data by ANN in continuous blending
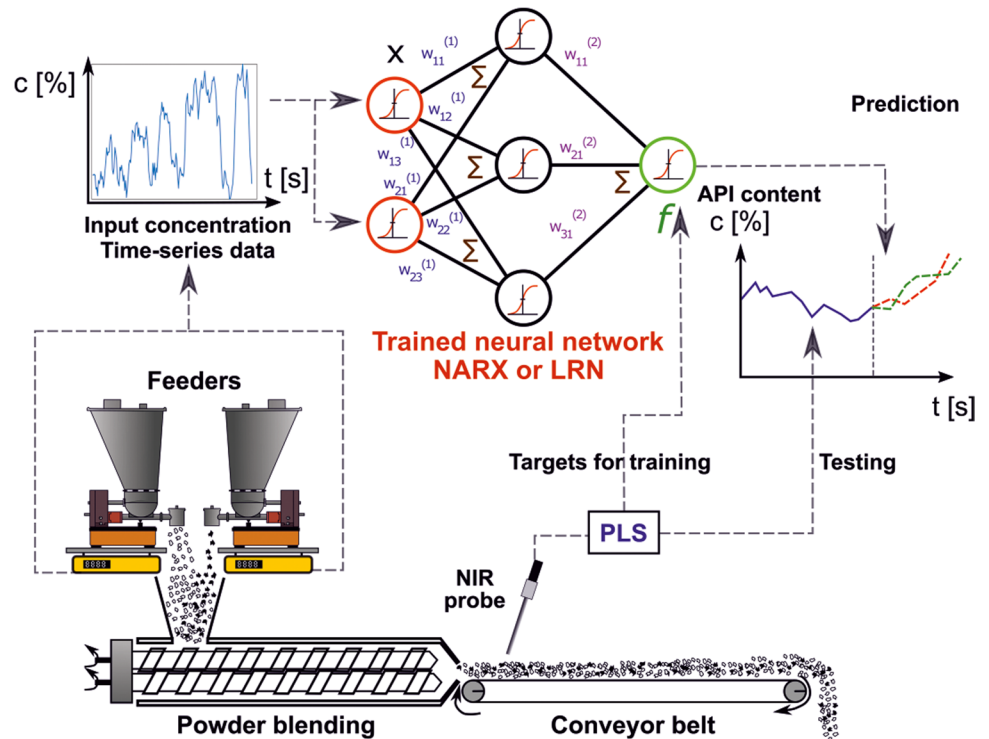
Granulation

Granulation is a particle enlargement technique that is often quintessential to ensure further processability, which can also greatly influence the quality (*e.g.*, content uniformity and dissolution) of the final product. Granulation is implemented as a wet or dry technique, *e.g.*, in high-shear, fluidized bed, roller compactor apparatus, or the continuous solution of twin-screw wet granulation (TSWG) is also emerging.

For more than 25 years, ANN models have been created to predict the quality of the product based on process parameters of fluidized bed (58–61), high-shear wet (62), and dry granulation (63, 64). For example, Kesavan *et al.* (65) modeled both high-shear and fluid granulation by ANNs to predict the particle size, flow rate, bulk density, and tap density. The inputs were the type and percentage of diluent, the type of granulation equipment, and the amount and addition method of the binder. The product CQAs could be predicted with good accuracy, and the ANN performed better than multilinear stepwise regression analysis. The process parameters have also been applied to predict the disintegration time of tablets compressed from the granules (66). Furthermore, the scale-up of wet and fluid bed granulation processes was also facilitated by ANNs (67–69). Korteby *et al.* (70) demonstrated with a fluid hot-melt granulation process that the relative importance of the independent input variables of the ANN model could be determined when combined with the Garson equation (71). They identified that the particle size

of the binder had the highest impact on the properties of the final granules, followed by binder viscosity grade and binder content. In this way, the ANN combined the advantages of first-principles and data-driven modeling by providing information about the effect of factors, while its construction was significantly easier than that of a first-principles model. In the case of dry granulation, the granule size distribution obtained after milling (72), the ribbon friability (63), or the ribbon density (73) could also be predicted. Modeling the granule quality based on operating parameters by ANN was also possible for continuous granulation (74), when the d_{10}, d_{50}, and d_{90} values of granules were calculated based on the liquid to solid ratio, screw speed, screw configuration, and material throughput. It has been suggested (74, 75) that the ANN models could be applied for the MPC of the process. Furthermore, ANNs can be combined with other data processing techniques such as Kriging of finite volume scheme to create hybrid models which ideally combine the benefits of both methods (76, 77), and consequently, ANN can be integrated into more complex systems, too.

AI can also process the data yielded by real-time sensors used as PAT tools in the granulation process, *e.g.*, to monitor the API content or the residual moisture content. Zhao *et al.* (78) measured the concentration of three APIs with an off-line NIR spectrometer in sugar-free Yangwei granules manufactured on a commercial-scale apparatus. The BP ANN and other ML methods yielded similar results to the PLS regression. Rantanen *et al.* (79) created PLS and ANN models to predict the moisture content of granules based on

NIR spectra. They found that the ANN had more predictive power for independent test samples. Gupta *et al.* (80) used a NIR and a microwave spectrometer to record spectra of a ribbon leaving a roll compactor. The pretreated spectra were processed by PLS and ANN models to predict the API content, moisture content, and density of the ribbons, where ANN and PLS performed similarly except for the moisture content.

The flexibility of ANNs allows us to process any kind of signal effectively. For instance, thermocouples could also serve as PAT tools. Korteby *et al.* (81) placed three sets of thermocouples inside a conical fluidized bed granulator. They recorded the temperatures under different conditions, as the temperature distribution inside the granulator can influence the granule properties; thus, understanding its dependence on manufacturing parameters can contribute to the creation of a more reliable process. The obtained data were used to train an ANN, which provided very accurate predictions for the test cases, and such a model could be the basis of a real-time quality control scheme. The acoustic emission of the fluidized bed granulation could also be monitored, as demonstrated by Carter *et al.* (82). They placed piezoelectric microphones in different positions outside the apparatus and recorded their signal while intentionally blocking parts of the distributor plate. After extracting time and frequency domain feature vectors from the sound signal, an ANN was trained to recognize different blockage scenarios based on the emitted sound. According to this revelation, AI can facilitate novel applications of acoustic signals, possibly leading to the creation of powerful PAT tools.

Tableting, Coating

In most pharmaceutical manufacturing processes, tableting creates the individual units of the end product. Ensuring that each tablet the patient receives meets the strict quality requirements is essential. The advent of predictive modeling and PAT technologies offers great help in achieving this goal.

One of the first things to consider when developing a tableting process is how the compressed powder mixture behaves inside the tablet press. The flowability of the blend must be good enough that each time the die is filled, an almost identical mass of powder moves into it. Kachrimanis *et al.* (83) used an FF-BP ANN to predict the flow rate of various powders through a circular orifice. They used typical powder properties as input, such as bulk density, tapped density, particle diameter, aspect ratio, roundness, convexity, and true density. The obtained flow rate predictions were more accurate than those of the flow equation proposed by Jones and Pilpel. Powder properties can also be utilized to predict the compressibility of the material. CMAs, such as the type and particle size of diluent, the type of glidant,

bulk density, Carr's compressibility index, and parameters of Kawakita's equation (84–86), were used with various ML algorithms based on the results of a design of experiments (DoE) consisting of 30–50 settings.

Capping, *i.e.*, the premature detachment of the tablet's top layers, means serious quality problem in further processing (*e.g.*, film coating and packaging) and should be avoided. Therefore, Belič *et al.* (87) predicted the capping tendency with neural networks and fuzzy logic by accounting for the particle size of the tableted powder and tablet press settings. They concluded that the technique makes formulation development significantly more effective than traditional trial-and-error approaches.

As a dosage form is developed, large datasets are created that enable the fitting of design spaces within the QbD approach by using suitable mathematical tools. Zawbaa *et al.* (88) applied a combination of ANN with variable selection algorithms to find which manufacturing parameters have the strongest influence on tablets' porosity and tensile strength. Results from the variable selection enabled the authors to identify that the compaction pressure was the dominant factor.

These studies showed that ANNs are suitable to determine the design space and predicting the processability of the powder and the quality of the tablets based on CMAs, but the tableting step is still lacking PAT-based ANN model applications. The CQA of the final tablets is influenced not only by the tableting step but also by the previous manufacturing techniques and the material attributes of the raw materials. Therefore, the works tackling the characterization of the final tablets are further detailed in the next section.

CHARACTERIZATION OF THE FINAL PRODUCTS

Content Uniformity, Assay

The content uniformity (CU) of the final products or intermediates is one of the most frequently studied CQAs that must fall within certain limits. Spectroscopic PAT tools are widely used to quantify the API content in solid dosage forms to reach these goals. However, linear quantitative methods are not always feasible for evaluating multivariate data. In these cases, ANNs may provide a solution to reach a validated CU method.

Traditionally, UV-Vis spectroscopy is used to analyze assays, and ANNs have been applied several times to improve the quantification of numerous APIs (89–91), even in minor amounts. However, it is a destructive technique which is not compatible with the PAT concept.

In contrast, vibrational spectroscopy, *e.g.*, Raman and NIR spectroscopy, can be helpful as an in-line, nondestructive

method for the characterization of solid samples. However, only one study has been found for the quantification of API by Raman spectroscopy and ANN (92), wherein commercial tablets and capsules containing diclofenac sodium were studied. PLS, principal component regression (PCR), and counter-propagation ANNs (CP-ANN) methods were compared, the latter combining unsupervised and supervised learning. While PCR yielded consistently higher errors, the PLS and CP-ANN showed comparable results for both tablets and capsules. A 2.6–3.5% and 1.4–1.7% relative standard error of validation was reached for tablets and capsules, respectively, and a good correlation with reference analysis was obtained for commercial formulations. NIR spectroscopy is a more widely used technique, despite the severe overlapping between the signals of the components. Several APIs have been studied by ANNs, such as paracetamol, caffeine, ciprofloxacin, aspirin, and phenacetin (53, 93, 94). Different variable selection techniques have also been tested for improving the quantification by NNs. For example, variable selection by orthogonal projection to latent structures (O-PLS), combined with ANN (95), genetic algorithms (GA-ANN) (96), or wavelet transformation (WT) (97), could be applied to increase model accuracy. WT could also be used for the dimensional reduction of the original spectra (98), which is an essential step in ANN building to decrease the computational demand of the training.

It can be concluded that for the API content determination in solid samples, ANNs mainly improved results compared to the linear multivariate methods, *e.g.*, PLS regression with the same number of calibration samples. Another emerging application of ANN might be predicting the amount of analyte from process data, even without spectroscopic measurements, which is the realization of the RTRT concept. For example, it was possible to estimate the ascorbic acid concentration in nutraceutical products from physicochemical properties, namely, pH, specific gravity, and viscosity (99). In this case, the ANN, which served as the soft sensor, provided a regression coefficient of 0.92 for the quantification.

Tensile strength, Friability

The appropriate hardness is also a CQA of the tablets, impacting the further processability, *e.g.*, the coating and the packaging, and is mainly characterized by the tensile strength (TS) or the friability (FR). However, these properties are not easily measurable by available PAT tools. Attempts have been made to monitor the TS by NIR spectroscopy, where, *e.g.*, the change of the baseline can correlate with the tablet hardness, which could be easily turned into a real-time technique (100, 101). By creating the optimal WT-ANN architecture, the tablet hardness was approximated satisfactorily, exceeding the accuracy of the linear PLS regression model. In another study (102), the

hardness of theophylline tablets was predicted similarly by PLS and ANN at the lowest set point, but ANN produced better results for harder tablets.

Another possible approach is modeling the TS and FR based on their CMAs and CPPs. An ANN network was shown by Bourquin (103), where the weight ratio of four ingredients, the dwell time, and the compression force were used as input, and the TS and FR were predicted as outputs. The predicted TS had a good correlation with the observed values (R^2=0.753), but for the friability, the ANN model gave a very slight correlation (R^2=0.413). In this case, the tendency of overfitting can be recognized, which could have been avoided by using a larger training set.

Similarly, an ensemble ANN was used to study the effect of the type and amount of the filler (*e.g.*, microcrystalline cellulose, HPMC, crospovidone/PVP) and lubricant (magnesium stearate, sodium stearyl fumarate), with different APIs (104, 105). In (105), the crushing strength could be predicted with below 0.1 N error in the range of 30–60 N.

Furthermore, tablet properties and tableting process parameters could also be incorporated into the NNs, such as diameter, compression force, weight, height, porosity, speed of sound in the radial direction, and tablet compression speed (106). In (107), the type of polymers and their concentration were varied to predict the tablets' tensile strength, the total work of compression, the detachment work, and the ejection work with six different ML algorithms, involving four ANN methods.

In Vitro Dissolution

In vitro dissolution testing is an important indicator of product quality and therefore plays a vital role in the research, development, and routine quality control of the drug products. The tests, however, need to be carried out in standardized instruments and are labor- and time-intensive methods. Consequently, it could greatly benefit from an RTRT approach, for which ANNs have also been studied.

However, most of the ANN studies connected to the prediction of the dissolution deal mostly with formulation optimization for the required dissolution properties. In this context, several different ANN structures have proved to be applicable, such as MLP, Elman networks, and RNNs (108–112), and several process parameters were modeled, such as the effect of retardation polymer in the tablets (108), the tableting compression force (108, 113), and the crushing strength (114).

PAT tools can be used for predicting the dissolution if the effect of the CMAs/CPPs on the dissolution is detectable in the PAT data. For example, NIR spectra with PLS regression predicted the dissolution where the variation of the moisture content (115), compression force (116), mixing shear forces (117), or tablet composition were the critical

factors (118). Pawar *et al.* used at-line NIR spectroscopy in a continuous direct compression process, where the API content, compression force, feed frame speed, and blender speed simultaneously influenced the dissolution (117). The use of Raman chemical maps to non-destructively predict the dissolution has also been recently demonstrated (119). In this case, not only the chemical composition of the tablets, but the spatial distribution and CSD of the components could also be derived from a chemical map. However, to use it as a PAT technique, the speed of the chemical mapping still need to be further decreased.

Applying a single PAT tool might not always be sufficient. ANNs can aid the data fusion of different PAT sources and process data for a surrogate dissolution model. Our group demonstrated the merging Raman and NIR spectra of an extended-release tablet formulation by an ANN first. The data-fused ANN models outperformed both the PLS modeling results, as well as the models built by only using a single PAT sensor (120). ANNs could be developed not only using spectroscopic data but including additional process data in the ANN (Fig. 3), such as the registered compression force (121) and CSD data (122). Furthermore, SVM and an ensemble of regression trees were also tested, but ANNs provided the most accurate results. The concept can be generalized for arbitrary numbers and types of input data, which could significantly aid the implementation of predictive dissolution models in an RTRT framework.

FUTURE PROSPECTS

By reviewing the existing ANN applications for the pharmaceutical manufacturing steps, we could identify two major groups of the works: the utilization of ANNs (1) for non-linear regression for the evaluation of analytical sensor data and (2) to establish a relationship between arbitrary input and output parameters. Table I summarizes the works where the developed models were based on PAT data or where the input could be directly collected during a process.

As for the first group, mainly UV and (N)IR spectra were applied, but Raman spectroscopy can also be identified as a rapidly emerging tool. ANNs were consistently recognized as comparable and often superior to traditional PLS regression. Comparable results are expected when no significant non-linear relationship exists between the inputs and outputs, while ANNs could be superior when there is strong non-linearity. ANNs might also provide inferior results. One of the possible reasons for this is the overfitting of the ANN model when there is not enough training data to adequately capture the input-output relationship. Furthermore, outlier training data can significantly diminish the predictive power of the model, as in this case, the network might be fitted to inadequate non-linear behavior. However, these problems can be eliminated by expanding the training dataset and applying outlier filtering techniques.
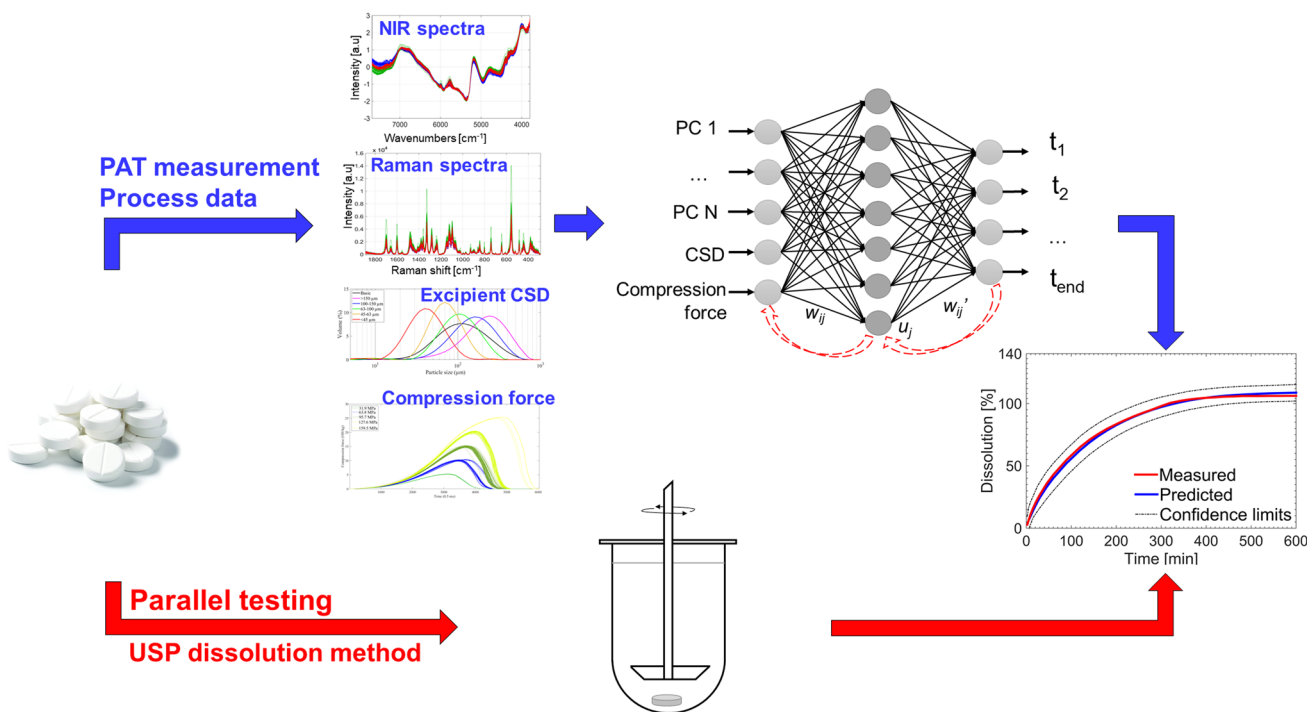

Fig. 3 Prediction of *in vitro* dissolution by neural network from PAT data

Table I Application of Neural Networks for (Potential) PAT Purposes in the Process Steps of Pharmaceutical Manufacturing

Studied process/ product	(Potential) PAT data used in ANN	Dataset	ANN type	Predicted output/purpose of ANN	Ref.
API synthesis					
Glucose fermentation	FT-IR	15 fermentation samples	FF-BP	Glucose and glucuronic acid concentration	(37)
Oxidative coupling of phenols	UV-Vis	16 samples	PC-ANN	Kinetic rate constant and reaction order	(38)
MPC of reversible reaction, continuous process	Concentration	Simulated time-series data	RNN	Model predictive control of concentration	(35)
Reaction in CSTR	Operation conditions	Time-series data	Elman NN	Concentration control	(36)
Crystallization					
Reactive crystallization of cephalexin	PVM	~2000–4000 images	ResNet CNN	Contamination classification	(43)
L-glutamic acid crystallization	High-resolution camera	160 images (>2000 crystals)	Mask R-CNN	*In situ* growth rate monitoring	(44)
Sucrose crystallization	FBRM, conc.	116 experiments	LRNN	Real-time CSD from FBRM data	(45)
KH_2PO_4 crystallization (computational study)	(FBRM, PVM)	Simulated CSDs	FF-BP	CLD, ARD (FBRM and PVM software sensor) from simulated CSD	(46)
Penicillin G Potassium crystallization	Temperature, flow rates	In-process data	Diagonal RNN	PID gains, self-tuning PID control	(47)
Batch crystallization	Conc., temperature	Simulation data from PBM	FF-BP	Future process response, temperature profile optimization	(48)
Industrial sugar crystallization	Feed mass flow rate	Industrial data from 6 months of campaign	Closed-loop ANN	Crystal mass prediction, NMPC	(49)
KCl cooling MSMPR crystallization	Flow rates, temperature, fines dissolution rate	2000 data points	RNN, FF-BP	MPC of crystal size, purity, production rate	(50, 51)
Ibuprofen batch crystallization	Temperature, FBRM	Reference batch data, real-time data	RBF	Control of CLD, real-time training	(52)
Powder blending					
Compound powdered drug	NIR	32 calibration samples	FF-BP	Paracetamol, diphenhydramine conc. prediction	(53)
Phenoxymethylpenicillin powder	NIR	66 calibration samples	FF-BP	API content prediction	(54)
4-component powder	NIR	111 calibration samples	FF-BP	API content prediction for blending endpoint detection	(55)
Caffeine–dextrose continuous blending	Feeder scale data	In-process data	NARX, LRN	API content prediction, at the blender output	(57)
Granulation					
Acetaminophen-MCC blend roller compaction	Microwave spectroscopy, NIR	Compacted ribbon samples, various settings	FF-BP	Prediction of granule density, moisture content, API content	(80)
Dry granulation of MCC	Roll force, screw speed	16 experiments	FF-BP	Prediction of ribbon density	(73)
Dry granulation of MCC–mannitol mixtures	Compaction force, roll speed	161 experiments	FF-BP	Prediction of granule size distribution	(72)
Fluid bed granulation of sugar and glass sphere pellets	Microphone	36 experiments	FF-BP	Detection of distributor plate blockage	(82)

Table I (continued)

Studied process/ product	(Potential) PAT data used in ANN	Dataset	ANN type	Predicted output/purpose of ANN	Ref.
Fluid bed granulation of theophylline	NIR	148 samples	FF-BP	Prediction of moisture content	(79)
Commercial-scale fluid bed granulation of *Yangwei*	NIR	268 samples	FF-BP	Prediction of API content	(78)
TSWG of MCC	Screw speed, powder flow rate	24 experiments + Kriging interpolation	FF-BP	Prediction of granule particle size, mean residence time	(74, 76)
Tableting, coating					
Tableting	Compression force	Tablets produced with 209 settings	FF-BP with optimization	Prediction of tablet porosity and tensile strength	(88)
Integrated manufacturing line					
Diclofenac sodium granulation in integrated continuous process (feeders, blender, TSWG, fluid bed drying, tableting film coating)	7 CPPs (API, excipient mass flows, liquid feed rate, rotation speeds, temperature, airflow)	148.000 data points recorded with 1 s intervals	FF-BP	Prediction of 8 CQAs: API content in 4 positions, LOD after dryer, CSD (d10 d50 d90) after dryer	(75)
High drug load IR formulation (high shear granulation, drying and tableting)	Process parameters of granulator, dryer, and tablet press	Batches manufactured over 3 years	FF-BP	Prediction of granulation water amount, tableting speed, disintegration time of cores	(66)
Content uniformity, assay					
5-component industrially formulated tablets with low dosage API	NIR	12*10 samples	FF-BP	Active content determination in tablets	(98)
Aspirin tablets	NIR	100 training samples	PC-ANN	Simultaneous determination of aspirin and phenacetin concentration	(94)
Ciprofloxacin powder	sw-NIR	50 training samples	BP	Concentration determination	(93)
Cefalexin tablets	sw-NIR	72 training samples	BP	Cefalexin determination	(96)
Diclofenac capsules/tablets	FT-Raman/NIR	23/80 training samples	CP, BP	Diclofenac determination	(92, 95)
Paracetamol tablets	NIR	30 training samples	FF-BP	Paracetamol content determination in tablet	(97)
Nutraceutical syrup production	Ascorbic acid conc., pH, specific gravity, and viscosity	70% of 264 batches	FF-BP	RTRT of ascorbic acid concentration	(99)
Tensile strength, friability					
Metformin tablet	NIR	66 training samples	BP reverse	Tablet hardness prediction	(100)
Naproxen tablet	NIR	60 training samples	FF-BP	Tablet hardness prediction	(101)
Theophylline tablet	NIR	4(*10) training samples	BP	Tablet hardness prediction	(102)
Galenical study on a tablet	Composition, dwell time, compression force	28 training samples	FF-BP	Tensile strength and friability prediction	(103)
Ramipril tablet	Filler/lubricant type and rate	32 training samples	BP	Crushing strength prediction	(105)

Table I (continued)

Studied process/ product	(Potential) PAT data used in ANN	Dataset	ANN type	Predicted output/purpose of ANN	Ref.
In vitro dissolution					
Caffeine tablet	NIR, Raman	32(*5) training samples	FF-BP	Full dissolution curve prediction	(120)
Drotaverine tablet	NIR, Raman, compression force, CSD	37–42 (*4) training samples	FF-BP, SVM, regression trees	Full dissolution curve prediction	(121, 122)

LRNN layer recurrent neural network, *NARX* non-linear autoregressive network with exogenous inputs, *PC-ANN* principal component-ANN, *FF-BP* feedforward backpropagation, *CP* counter-propagation, *(N)MPC* non-linear model predictive control, *MSMPR* mixed suspension-mixed product removal (crystallizer); *sw-NIR* short-wavelength NIR, *LOD* limit of detection, *MCC* microcrystalline cellulose

Therefore, it might be worth considering the application of ANNs for spectral data evaluation when PLS models cannot provide sufficient accuracy due to a possible non-linear effect. It is also worth noting that the development of ANNs did not require significantly more calibration data than for the PLS models, which is a common preconception about ANNs. Different types of ANNs have been used in this context with varying dimensional reduction methods, such as using principal component scores or wavelet transformation. However, to the best of the authors' knowledge, these techniques have not been thoroughly compared yet. Consequently, their systematic evaluation could significantly contribute to establishing good modeling practices to facilitate the application of ANNs. It is also noticeable that deep learning with spectroscopic data has rarely been applied. However, in (123), Zhang *et al.* demonstrated by using four different IR datasets (corn, wheat, soil, and pharmaceutical tablets) that deep NNs provide improved accuracy compared to conventional quantitative analysis. The same conclusions were drawn with other deep NNs with agricultural and food IR data (124, 125).

Although spectroscopies are probably the most common PAT tools, ANNs could be successfully applied for other types of PAT sensors, such as acoustic emission (82), FBRM, image analysis (44, 45), or for fusing different PAT data (*e.g.*, NIR and Raman) (120). The application of ML might be especially important for image analysis, for which deep learning has already proved its capacity, but mainly outside the pharmaceutical manufacturing (*e.g.*, in self-driving cars). Therefore, it would be worth further investigating NNs for machine vision in the pharmaceutical industry (126), *e.g.*, to identify faulty tablets automatically.

Most works have been found related to establishing relationships between CMAs, CPPs, and different CQAs of the individual process steps. These studies demonstrate the capability of ANNs utilized within the QbD framework to fit, *e.g.*, design spaces, primarily using a limited number of (off-line) designed experiments (usually 30–70 experiments) or historical data. However, the most significant shortcoming of this approach is that it does not realize the PAT initiative. Nevertheless, possibly the biggest prospect of ANNs lies in incorporating this approach into the PAT concept by utilizing these models in real-time, using the actual material and process parameters and the *in situ* registered PAT measurements. Furthermore, as the previous sections demonstrate, the individual unit operations have already been explored by several researchers, but ANNs have rarely been incorporated into integrated processes, yet. Roggo *et al.* (75) reported on its first realization by examining a manufacturing line consisting of feeding, TSWG, fluid bed drying, sieving, and tableting. Seven CPPs were recorded with a frequency of 1 s for a total of 148.000 data points, which were used to predict eight different CQAs of intermediate and end products.

The developed deep NNs (3 hidden layers) could learn from noisy PAT data and, consequently, be utilized for real-time control of continuous systems. Further studies are still needed in the future to examine the capabilities of machine learning in integrated manufacturing processes, which is the ultimate aim of commercializing the concept.

Digital Transformation

Following the Pharma 4.0 concept, digitalization is expected to spread significantly in the following years, as it can considerably improve the transparency, flexibility, efficiency, productivity, and quality of manufacturing (127). The authors of (128) from Novartis Global Drug Development—a leader in the digitalization of the pharmaceutical industry—have expressed that the historical operational data could be the goldmine to represent the pharmaceutical company's experience. However, this information is currently greatly fragmented, inconsistent, and time-consuming to reach. Digitalization platforms, such as the "Nerve Live" platform of Novartis (128), could help collect, clean, and analyze this goldmine. For example, centralized, easily accessible databases (data lakes) of raw materials' attributes, process parameters of each unit operation and the different PAT measurements could be obtained for the manufacturing process, as illustrated in Fig. 4.

Digitalization imposes multiple challenges on pharmaceutical companies and initializes changes at the business, operational, and technological levels (129). First of all, the role of data scientists and informational technology (IT) personnel significantly grows as new competencies and resources are required. For example, it is necessary to build cross-functional teams, cybersecurity needs to be addressed, and in the long term, standardization will be essential to assure compatibility (127). Multiple handbook chapters deal with the digital transformation of laboratories (*e.g.*, analytical, research, solid-state labs) (130, 131), providing an information knowledge base to the central concepts and providing guidance for their practical realization. For example, information management tools, *e.g.*, the Electronic Laboratory Notebook (ELN), Laboratory Information Management System (LIMS), and Enterprise Resource Planning (ERP), are introduced, and principles of cybersecurity, communication protocols, data and modeling technologies, reporting, and creating FAIR (findable, accessible, interoperable, reusable) data are discussed.

Most of these concepts are extensible to the operation of the labs and manufacturing sites of pharmaceutical companies, too. In (132), a complex IT infrastructure is proposed explicitly for pharmaceutical development, involving data management tools, knowledge modeling, and information sharing guidelines to aid the managing and interpreting different sources of complex information. The industrial realization of such an IT system is the "Nerve Live" platform of Novartis (128) or the commercially available cloud-based, open IoT operating system of Siemens, called MindSphere (129). MindSphere integrates all data sources and information management tools, connects different types of equipment in a cloud platform, and provides the highest level of data protection and storage, which is crucial for pharmaceutical companies.

Fig. 4 Artificial intelligence models for PAT in the Pharma 4.0 concept
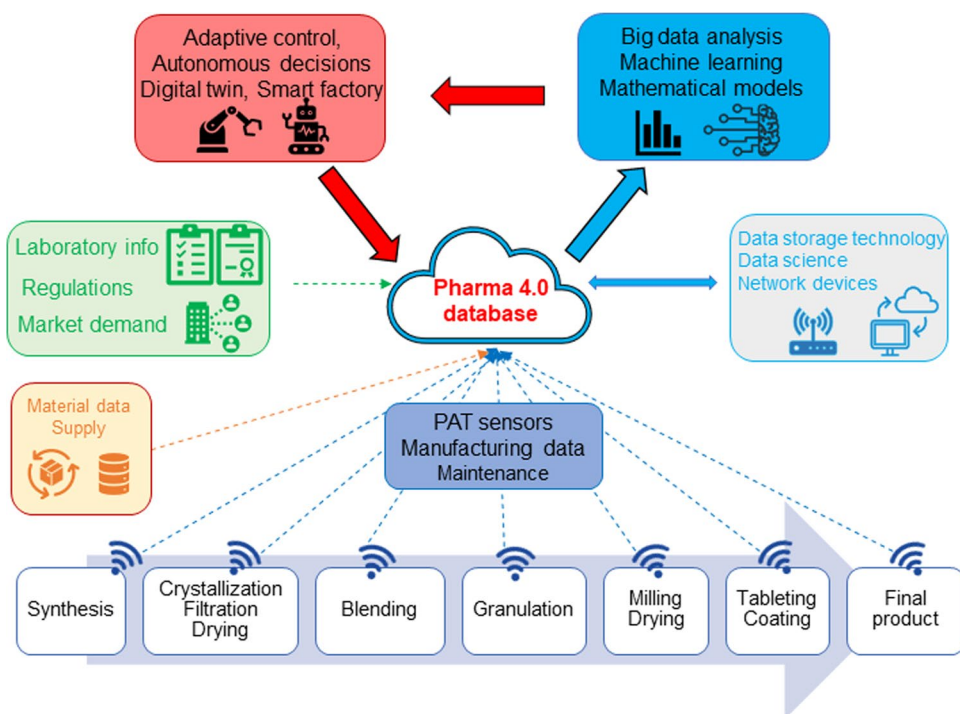
In all publications related to digitalization, the central role of AI and ML is recognized to analyze the data lake, AI solutions standing above the whole hierarchy of data (Fig. 4.), having access to each data management level (130). In this way, NN models can be built to monitor CQAs of processes, develop digital twins, and realize model-predictive control for intelligent decisions, self-optimization, predictive maintenance, or making business decisions. Several software solutions are aiming to realize this, *e.g.*, the Siemens Predictive Analysis (SiePA) tool is readily available to optimize a process based on historical data, analyze failure modes, and realize predictive maintenance (133). Despite, based on the published scientific papers, we could conclude that the application of ANN with real-time collected pharmaceutical manufacturing data and integrated multiple process steps is still scarce. However, it is worth noting that implementing such platforms means a substantial competitive advantage for pharmaceutical companies, which might cause the publication of these results to be hindered.

Nevertheless, we believe that further academic and industrial research would significantly facilitate the widespread realization of digital transformation and autonomous smart factories. For example, the adaptation and real-time training capabilities of NNs for continually increasing data should be further studied, as well as the role of time-series ANNs could be investigated in more detail. Furthermore, it is vital that an AI black-box model should not replace and diminish the scientific, chemical understanding of the research, development, and manufacturing process of drug products, which could be achieved by integrating physical-chemical process knowledge into the automatized platform, *e.g.*, as a form of hybrid mathematical models.

CONCLUSIONS

ML techniques, such as ANNs, have emerged as one of the essential data analysis tools for processing big data and realizing the Industry/Pharma 4.0 concepts. To assess the readiness of pharmaceutical manufacturing for this, this paper aimed to review the application of ANNs in the context of PAT. It can be concluded that ANNs have already been tested for several purposes in the most common manufacturing steps, but their real-time application for PAT is still scarce. The possible future directions and research gaps have also been identified. In this way, ANNs could significantly contribute to realizing smart, autonomous pharmaceutical manufacturing lines in the future. This can help the faster, cost-effective production or the reduction of waste, *i.e.*, reducing the environmental load, and the automatized systems could minimize human exposure to dangerous processes or drug substances, *e.g.*, hormones or cytostatics.

Declarations

References

1. Xu LD, Xu EL, Li L. Industry 4.0: state of the art and future trends. Int J Prod Res. 2018;56(8):2941–62.
2. Arden NS, Fisher AC, Tyner K, Yu LX, Lee SL, Kopcha M. Industry 4.0 for pharmaceutical manufacturing: preparing for the smart factories of the future. Int J Pharm. 2021;602:120554.
3. Barenji RV, Akdag Y, Yet B, Oner L. Cyber-physical-based PAT (CPb-PAT) framework for Pharma 4.0. Int J Pharm. 2019;567:118445.
4. Kusiak A. Smart manufacturing. Int J Prod Res. 2018;56(1-2):508–17.
5. The international conference on harmonization of technical requirements for registration of pharmaceuticals for human use (ICH), Quality Guideline Q8 Pharmaceutical Development. 2009.
6. FDA. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry: PAT – a framework for innovative pharmaceutical development, manufacturing, and quality assurance. 2004.
7. EMA. European Medical Agency, Guideline on real time release testing (formerly Guideline on parametric release). 2012.
8. Wuest T, Weimer D, Irgens C, Thoben K-D. Machine learning in manufacturing: advantages, challenges, and applications. Prod Manuf Res. 2016;4(1):23–45.
9. Ren S, Zhang Y, Liu Y, Sakao T, Huisingh D, Almeida CMVB. A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: a framework, challenges and future research directions. J Clean Prod. 2019;210:1343–65.
10. Alexopoulos K, Nikolakis N, Chryssolouris G. Digital twin-driven supervised machine learning for the development of

artificial intelligence applications in manufacturing. Int J Comput Integr Manuf. 2020;33(5):429–39.

11. Chen Y, Yang O, Sampat C, Bhalode P, Ramachandran R, Ierapetritou M. Digital twins in pharmaceutical and biopharmaceutical manufacturing: a literature review. Processes. 2020;8(9):1088.

12. Joshi AV. Introduction to AI and ML. In: Joshi AV, editor. Machine learning and artificial intelligence. Cham: Springer International Publishing; 2020. p. 3–7.

13. Danish Medicines Agency (DKMA). Suggested criteria for using AI/ML algorithms in GxP. 2021. https://laegemiddelstyrelsen.dk/en/licensing/supervision-and-inspection/inspection-of-authorised-pharmaceutical-companies/using-aiml-algorithms-in-gxp/. Accessed 9 Feb 2022.

14. FDA. U.S. Center for devices and radiological health, Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. 2021. https://www.fda.gov/media/145022/download. Accessed 9 Feb 2022.

15. Mitchell JBO. Machine learning methods in chemoinformatics. WIREs Comput Mol Sci. 2014;4(5):468–81.

16. Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal. 2000;22(5):717–27.

17. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463–77.

18. Réda C, Kaufmann E, Delahaye-Duriez A. Machine learning applications in drug development. Comput Struct Biotechnol J. 2020;18:241–52.

19. Puri M, Solanki A, Padawer T, Tipparaju SM, Moreno WA, Pathak Y. Chapter 1 - Introduction to Artificial Neural Network (ANN) as a predictive tool for drug design, discovery, delivery, and disposition: basic concepts and modeling. In: Puri M, Pathak Y, Sutariya VK, Tipparaju S, Moreno W, editors. Artificial neural network for drug design, delivery and disposition. Boston: Academic Press; 2016. p. 3–13.

20. Đuriš J, Kurćubić I, Ibrić S. Review of machine learning algorithms application in pharmaceutical technology. Arhiv za farmaciju. 2021;71(4):302–17.

21. Lou H, Lian B, Hageman MJ. Applications of machine learning in solid oral dosage form development. J Pharm Sci. 2021;110(9):3150–65.

22. Salehinejad H, Sankar S, Barfett J, Colak E, Valaee S. Recent advances in recurrent neural networks. arXiv preprint arXiv:180101078. 2017.

23. Hornik K. Approximation capabilities of multilayer feedforward networks. Neural Netw. 1991;4(2):251–7.

24. Dhillon A, Verma GK. Convolutional neural network: a review of models, methodologies and applications to object detection. Prog Artif Intell. 2020;9(2):85–112.

25. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8(1):53.

26. Mittal A, Singh AP, Chandra P. A modification to the Nguyen–Widrow weight initialization method. In: Thampi SM, Trajkovic L, Mitra S, Nagabhushan P, Mukhopadhyay J, Corchado JM, et al., editors. Intelligent systems, technologies and applications. Singapore: Springer Singapore; 2020. p. 141–53.

27. Battiti R. First- and Second-order methods for learning: between steepest descent and Newton's method. Neural Comput. 1992;4(2):141–66.

28. Hagan MT, Menhaj MB. Training feedforward networks with the Marquardt algorithm. IEEE Trans Neural Netw. 1994;5(6):989–93.

29. MacKay DJ. Bayesian interpolation. Neural Comput. 1992;4(3):415–47.

30. Foresee FD, Hagan MT, editors. Gauss-Newton approximation to Bayesian learning. Proceedings of International Conference on Neural Networks (ICNN'97); 1997.

31. Moghaddam MG, Ahmad FBH, Basri M, Rahman MBA. Artificial neural network modeling studies to predict the yield of enzymatic synthesis of betulinic acid ester. Electron J Biotechnol. 2010;13:3–4.

32. Valizadeh H, Pourmahmood M, Mojarrad JS, Nemati M, Zakeri-Milani P. Application of artificial intelligent tools to modeling of glucosamine preparation from exoskeleton of shrimp. Drug Dev Ind Pharm. 2009;35(4):396–407.

33. Zhou Z, Li X, Zare RN. Optimizing chemical reactions with deep reinforcement learning. ACS Central Sci. 2017;3(12):1337–44.

34. Gbadago DQ, Moon J, Kim M, Hwang S. A unified framework for the mathematical modelling, predictive analysis, and optimization of reaction systems using computational fluid dynamics, deep neural network and genetic algorithm: a case of butadiene synthesis. Chem Eng J. 2021;409:128163.

35. Wong WC, Chee E, Li J, Wang X. Recurrent neural network-based model predictive control for continuous pharmaceutical manufacturing. Mathematics. 2018;6(11):242.

36. Baranilingesan I. Optimization algorithm based Elman neural network controller for continuous stirred tank reactor process model. Curr Sci. 2021;120(8):1324–33.

37. Franco VG, Perín JC, Mantovani VE, Goicoechea HC. Monitoring substrate and products in a bioprocess with FTIR spectroscopy coupled to artificial neural networks enhanced with a genetic-algorithm-based method for wavelength selection. Talanta. 2006;68(3):1005–12.

38. Hasani M, Moloudi M. Application of principal component-artificial neural network models for simultaneous determination of phenolic compounds by a kinetic spectrophotometric method. J Hazard Mater. 2008;157(1):161–9.

39. Gao Y, Zhang T, Ma Y, Xue F, Gao Z, Hou B, et al. Application of PAT-based feedback control approaches in pharmaceutical crystallization. Crystals. 2021;11(3):221.

40. Velásco-Mejía A, Vallejo-Becerra V, Chávez-Ramírez AU, Torres-González J, Reyes-Vidal Y, Castañeda-Zaldivar F. Modeling and optimization of a pharmaceutical crystallization process by using neural networks and genetic algorithms. Powder Technol. 2016;292:122–8.

41. Shaikh R, Shirazian S, Walker GM. Application of artificial neural network for prediction of particle size in pharmaceutical cocrystallization using mechanochemical synthesis. Neural Comput Appl. 2021;33(19):12621–40.

42. Vasanth Kumar K, Martins P, Rocha F. Modelling of the batch sucrose crystallization kinetics using artificial neural networks: comparison with conventional regression analysis. Ind Eng Chem Res. 2008;47(14):4917–23.

43. Salami H, McDonald MA, Bommarius AS, Rousseau RW, Grover MA. In situ imaging combined with deep learning for crystallization process monitoring: application to cephalexin production. Org Process Res Dev. 2021;25(7):1670–9.

44. Chen S, Liu T, Xu D, Huo Y, Yang Y, editors. Image based measurement of population growth rate for L-glutamic acid crystallization. 2019 Chinese Control Conference (CCC); 2019 27-30 July 2019.

45. Crestani CE, Bernardo A, Costa CBB, Giulietti M. An artificial neural network model applied to convert sucrose chord length distributions into particle size distributions. Powder Technol. 2021;384:186–94.

46. Szilágyi B, Nagy ZK. Aspect ratio distribution and chord length distribution driven modeling of crystallization of

two-dimensional crystals for real-time model-based applications. Cryst Growth Des. 2018;18(9):5311–21.

47. Song Z, Liu C, Song X, editors. Application of self-tuning Pid control based on diagonal recurrent neural network in crystallization process. 2006 International Conference on Machine Learning and Cybernetics; 2006 13-16 Aug. 2006.

48. Paengjuntuek W, Thanasinthana L, Arpornwichanop A. Neural network-based optimal control of a batch crystallizer. Neurocomputing. 2012;83:158–64.

49. Damour C, Benne M, Grondin-Perez B, Chabriat J-P. Nonlinear predictive control based on artificial neural network model for industrial crystallization. J Food Eng. 2010;99(2):225–31.

50. Rohani S, Haeri M, Wood HC. Modeling and control of a continuous crystallization process Part 1. Linear and non-linear modeling. Comput Chem Eng. 1999;23(3):263–77.

51. Rohani S, Haeri M, Wood HC. Modeling and control of a continuous crystallization process Part 2. Model predictive control. Comput Chem Eng. 1999;23(3):279–86.

52. Öner M, Montes FCC, Ståhlberg T, Stocks SM, Bajtner JE, Sin G. Comprehensive evaluation of a data driven control strategy: experimental application to a pharmaceutical crystallization process. Chem Eng Res Des. 2020;163:248–61.

53. Dou Y, Sun Y, Ren Y, Ren Y. Artificial neural network for simultaneous determination of two components of compound paracetamol and diphenhydramine hydrochloride powder on NIR spectroscopy. Anal Chim Acta. 2005;528(1):55–61.

54. Wang B, Liu G, Liu S, Fei Q, Ren Y. Orthogonal projection to latent structures combined with artificial neural network for quantitative analysis of phenoxymethylpenicillin potassium powder. Vib Spectrosc. 2009;51(2):199–204.

55. Tewari J, Strong R, Boulas P. At-line determination of pharmaceuticals small molecule's blending end point using chemometric modeling combined with Fourier transform near infrared spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc. 2017;173:886–91.

56. Mujumdar A, Robi PS, Malik M, Horio M. Artificial neural network (ANN) model for prediction of mixing behavior of granular flows. IJCEM. 2007;8(3):149–58.

57. Beke ÁK, Gyürkés M, Nagy ZK, Marosi G, Farkas A. Digital twin of low dosage continuous powder blending – artificial neural networks and residence time distribution models. Eur J Pharm Biopharm. 2021;169:64–77.

58. Petrovic JM, Chansanroj K, Meier B, Ibrj S, Betz G. Analysis of fluidized bed granulation process using conventional and novel modeling techniques. Eur J Pharm Sci. 2011;44(3):227–34.

59. Behzadi SS, Klocker J, Hüttlin H, Wolschann P, Viernstein H. Validation of fluid bed granulation utilizing artificial neural network. Int J Pharm. 2005;291(1):139–48.

60. Murtoniemi E, Yliruusi J, Kinnunen P, Merkku P, Leiviskä K. The advantages by the use of neural networks in modelling the fluidized bed granulation process. Int J Pharm. 1994;108(2):155–64.

61. Behzadi SS, Prakasvudhisarn C, Klocker J, Wolschann P, Viernstein H. Comparison between two types of artificial neural networks used for validation of pharmaceutical processes. Powder Technol. 2009;195(2):150–7.

62. Sampat C, Ramachandran R. Identification of granule growth regimes in high shear wet granulation processes using a physics-constrained neural network. Processes. 2021;9(5):737.

63. Inghelbrecht S, Remon J-P, Fernandes de Aguiar P, Walczak B, Massart D, Van De Velde F, et al. Instrumentation of a roll compactor and the evaluation of the parameter settings by neural networks. Int J Pharm. 1997;148(1):103–15.

64. Turkoglu M, Aydin I, Murray M, Sakr A. Modeling of a roller-compaction process using neural networks and genetic algorithms. Eur J Pharm Biopharm. 1999;48(3):239–45.

65. Kesavan JG, Peck GE. Pharmaceutical granulation and tablet formulation using neural networks. Pharm Dev Technol. 1996;1(4):391–404.

66. Mathe R, Casian T, Tomuţă I. Multivariate feed forward process control and optimization of an industrial, granulation based tablet manufacturing line using historical data. Int J Pharm. 2020;591:119988.

67. Watano S, Sato Y, Miyanami K. Application of a neural network to granulation scale-up. Powder Technol. 1997;90(2):153–9.

68. Millen N, Kovačević A, Khera L, Djuriš J, Ibric S. Machine learning modelling of wet granulation scale-up using compressibility, compactibility and manufacturability parameters. Hemijska Industrija. 2019;73(3):155–68.

69. Landin M. Artificial intelligence tools for scaling up of high shear wet granulation process. J Pharm Sci. 2017;106(1):273–7.

70. Korteby Y, Kristó K, Sovány T, Regdon G. Use of machine learning tool to elucidate and characterize the growth mechanism of an in-situ fluid bed melt granulation. Powder Technol. 2018;331:286–95.

71. Garson GD, editor. Interpreting neural-network connection weights 1991.

72. Kazemi P, Khalid MH, Pérez Gago A, Kleinebudde P, Jachowicz R, Szlęk J, et al. Effect of roll compaction on granule size distribution of microcrystalline cellulose-mannitol mixtures: computational intelligence modeling and parametric analysis. Drug Des Devel Ther. 2017;11:241–51.

73. Sajjia M, Shirazian S, Kelly CB, Albadarin AB, Walker G. ANN Analysis of a roller compaction process in the pharmaceutical industry. Chem Eng Technol. 2017;40(3):487–92.

74. Shirazian S, Kuhs M, Darwish S, Croker D, Walker GM. Artificial neural network modelling of continuous wet granulation using a twin-screw extruder. Int J Pharm. 2017;521(1):102–9.

75. Roggo Y, Jelsch M, Heger P, Ensslin S, Krumme M. Deep learning for continuous manufacturing of pharmaceutical solid dosage form. Eur J Pharm Biopharm. 2020;153:95–105.

76. Ismail HY, Singh M, Darwish S, Kuhs M, Shirazian S, Croker DM, et al. Developing ANN-Kriging hybrid model based on process parameters for prediction of mean residence time distribution in twin-screw wet granulation. Powder Technol. 2019;343:568–77.

77. Ismail HY, Singh M, Shirazian S, Albadarin AB, Walker GM. Development of high-performance hybrid ANN-finite volume scheme (ANN-FVS) for simulation of pharmaceutical continuous granulation. Chem Eng Res Des. 2020;163:320–6.

78. Zhao J, Tian G, Qiu Y, Qu H. Rapid quantification of active pharmaceutical ingredient for sugar-free Yangwei granules in commercial production using FT-NIR spectroscopy based on machine learning techniques. Spectrochim Acta A Mol Biomol Spectrosc. 2021;245:118878.

79. Rantanen J, Räsänen E, Antikainen O, Mannermaa J-P, Yliruusi J. In-line moisture measurement during granulation with a four-wavelength near-infrared sensor: an evaluation of process-related variables and a development of non-linear calibration model. Chemometr Intell Lab Syst. 2001;56(1):51–8.

80. Gupta A, Austin J, Davis S, Harris M, Reklaitis G. A novel microwave sensor for real-time online monitoring of roll compacts of pharmaceutical powders online - a comparative case study with NIR. J Pharm Sci. 2015;104(5):1787–94.

81. Korteby Y, Mahdi Y, Azizou A, Daoud K, Regdon G. Implementation of an artificial neural network as a PAT tool for the prediction of temperature distribution within a pharmaceutical fluidized bed granulator. Eur J Pharm Sci. 2016;88(Supplement C):219-32.

82. Carter A, Briens L. An application of deep learning to detect process upset during pharmaceutical manufacturing using passive acoustic emissions. Int J Pharm. 2018;552(1):235–40.

83. Kachrimanis K, Karamyan V, Malamataris S. Artificial neural networks (ANNs) and modeling of powder flow. Int J Pharm. 2003;250(1):13–23.

84. Barmpalexis P, Karagianni A, Karasavvaides G, Kachrimanis K. Comparison of multi-linear regression, particle swarm optimization artificial neural networks and genetic programming in the development of mini-tablets. Int J Pharm. 2018;551(1):166–76.

85. Khalid GM, Usman AG. Application of data-intelligence algorithms for modeling the compaction performance of new pharmaceutical excipients. Future J Pharm Sci. 2021;7(1):31.

86. Lou H, Chung JI, Kiang YH, Xiao L-Y, Hageman MJ. The application of machine learning algorithms in understanding the effect of core/shell technique on improving powder compactability. Int J Pharm. 2019;555:368–79.

87. Belič A, Škrjanc I, Božič DZ, Karba R, Vrečer F. Minimisation of the capping tendency by tableting process optimisation with the application of artificial neural networks and fuzzy models. Eur J Pharm Biopharm. 2009;73(1):172–8.

88. Zawbaa HM, Schiano S, Perez-Gandarillas L, Grosan C, Michrafy A, Wu C-Y. Computational intelligence modelling of pharmaceutical tabletting processes using bio-inspired optimization algorithms. Adv Powder Technol. 2018;29(12):2966–77.

89. Goicoechea HC, Collado MS, Satuf ML, Olivieri AC. Complementary use of partial least-squares and artificial neural networks for the non-linear spectrophotometric analysis of pharmaceutical samples. Anal Bioanal Chem. 2002;374(3):460–5.

90. Naguib IA, Darwish HW. Support vector regression and artificial neural network models for stability indicating analysis of mebeverine hydrochloride and sulpiride mixtures in pharmaceutical preparation: a comparative study. Spectrochim Acta A Mol Biomol Spectrosc. 2012;86:515–26.

91. Akbari Hasanjani HR, Sohrabi MR. Artificial neural networks (ANN) for the simultaneous spectrophotometric determination of fluoxetine and sertraline in pharmaceutical formulations and biological fluid. Iran J Pharm Res. 2017;16(2):478–89.

92. Mazurek S, Szostak R. Quantitative determination of diclofenac sodium in solid dosage forms by FT-Raman spectroscopy. J Pharm Biomed Anal. 2008;48(3):814–21.

93. Zhao LZ, Guo Y, Dou Y, Wang B, Mi H, Ren YL. Application of artificial neural networks to the nondestructive determination of ciprofloxacin hydrochloride in powder by short-wavelength NIR spectroscopy. J Anal Chem. 2007;62(12):1156–62.

94. Dou Y, Qu N, Wang B, Chi YZ, Ren YL. Simultaneous determination of two active components in compound aspirin tablets using principal component artificial neural networks (PC-ANNs) on NIR spectroscopy. Eur J Pharm Sci. 2007;32(3):193–9.

95. Wang B, Liu G, Dou Y, Liang L, Zhang H, Ren Y. Quantitative analysis of diclofenac sodium powder via near-infrared spectroscopy combined with artificial neural network. J Pharm Biomed Anal. 2009;50(2):158–63.

96. Huan Y, Feng G, Wang B, Ren Y, Fei Q. Quantitative analysis of cefalexin based on artificial neural networks combined with modified genetic algorithm using short near-infrared spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc. 2013;109:308–12.

97. Wang Z, Xiang B. Application of artificial neural network to determination of active principle ingredient in pharmaceutical quality control based on near infrared spectroscopy. Microchem J. 2008;89(1):52–7.

98. Chalus P, Walter S, Ulmschneider M. Combined wavelet transform–artificial neural network use in tablet active content determination by near-infrared spectroscopy. Anal Chim Acta. 2007;591(2):219–24.

99. Felipe MA, Baldovino R, editors. Real time release approach: at-line prediction of ascorbic acid concentration in nutraceutical syrup via artificial neural network2021; Cham: Springer International Publishing.

100. Wu J, Luo W, Wang X, Cheng Q, Sun C, Li H. A new application of WT-ANN method to control the preparation process of metformin hydrochloride tablets by near infrared spectroscopy compared to PLS. J Pharm Biomed Anal. 2013;80:186–91.

101. Luo W, Liu Y, Peng F, Li S, Li H. Enhanced characterization of naproxen formulation by near infrared spectroscopy. Anal Lett. 2014;47(14):2384–93.

102. Chen Y, Thosar SS, Forbess RA, Kemper MS, Rubinovitz RL, Shukla AJ. Prediction of drug content and hardness of intact tablets using artificial neural network and near-infrared spectroscopy. Drug Dev Ind Pharm. 2001;27(7):623–31.

103. Bourquin J, Schmidli H, van Hoogevest P, Leuenberger H. Comparison of artificial neural networks (ANN) with classical modelling techniques using different experimental designs and data from a galenical study on a solid dosage form. Eur J Pharm Sci. 1998;6(4):287–300.

104. Takagaki K, Arai H, Takayama K. Creation of a tablet database containing several active ingredients and prediction of their pharmaceutical characteristics based on ensemble artificial neural networks. J Pharm Sci. 2010;99(10):4201–14.

105. Aksu B, De Matas M, Cevher E, Özsoy Y, Güneri T, York P. Quality by design approach for tablet formulations containing spray coated ramipril by using artificial intelligence techniques. Int J Drug Deliv. 2012;4(1):11.

106. Akseli I, Xie J, Schultz L, Ladyzhynsky N, Bramante T, He X, et al. A Practical framework toward prediction of breaking force and disintegration of tablet formulations using machine learning tools. J Pharm Sci. 2017;106(1):234–47.

107. Djuris J, Cirin-Varadjan S, Aleksic I, Djuris M, Cvijic S, Ibric S. Application of machine-learning algorithms for better understanding of tableting properties of lactose co-processed with lipid excipients. Pharmaceutics. 2021;13(5):663.

108. Ibrić S, Djuriš J, Parojčić J, Djurić Z. Artificial neural networks in evaluation and optimization of modified release solid dosage forms. Pharmaceutics. 2012;4(4):531.

109. Goh WY, Lim CP, Peh KK, Subari K. Application of a recurrent neural network to prediction of drug dissolution profiles. Neural Comput Appl. 2002;10(4):311–7.

110. Sun Y, Peng Y, Chen Y, Shukla AJ. Application of artificial neural networks in the design of controlled release drug delivery systems. Adv Drug Del Rev. 2003;55(9):1201–15.

111. Leane MM, Cumming I, Corrigan OI. The use of artificial neural networks for the selection of the most appropriate formulation and processing variables in order to predict the in vitro dissolution of sustained release minitablets. AAPS PharmSciTech. 2003;4(2):129–40.

112. Mendyk A, Güres S, Jachowicz R, Szlwk J, Polak S, Wisniowska B, et al. From heuristic to mathematical modeling of drugs dissolution profiles: application of artificial neural networks and genetic programming. Comput Math Methods Med. 2015;2015:9.

113. Petrović J, Ibrić S, Betz G, Parojčić J, Đurić Z. Application of dynamic neural networks in the modeling of drug release from polyethylene oxide matrix tablets. Eur J Pharm Sci. 2009;38(2):172–80.

114. Ivic B, Ibric S, Betz G, Zorica D. Optimization of drug release from compressed multi unit particle system (MUPS) using generalized regression neural network (GRNN). Arch Pharm Res. 2010;33(1):103–13.

115. Zannikos PN, Li W-I, Drennen JK, Lodder RA. Spectrophotometric prediction of the dissolution rate of carbamazepine tablets. Pharm Res. 1991;8(8):974–8.

33333

312312312