Review of machine learning algorithms´ application in pharmaceutical technology

Jelena Djuris, Ivana Kurcubic, Svetlana Ibric

University of Belgrade – Faculty of Pharmacy, Department of Pharmaceutical Technology and Cosmetology, Vojvode Stepe 450, 11221 Belgrade, Serbia

Corresponding author: Jelena Djuris, e-mail: jelena.djuris@pharmacy.bg.ac.rs

Abstract

Machine learning algorithms, and artificial intelligence in general, have a wide range of applications in the field of pharmaceutical technology. Starting from the formulation development, through a great potential for integration within the Quality by design framework, these data science tools provide a better understanding of the pharmaceutical formulations and respective processing. Machine learning algorithms can be especially helpful with the analysis of the large volume of data generated by the Process analytical technologies. This paper provides a brief explanation of the artificial neural networks, as one of the most frequently used machine learning algorithms. The process of the network training and testing is described and accompanied with illustrative examples of machine learning tools applied in the context of pharmaceutical formulation development and related technologies, as well as an overview of the future trends. Recently published studies on more sophisticated methods, such as deep neural networks and light gradient boosting machine algorithm, have been described. The interested reader is also referred to several official documents (guidelines) that pave the way for a more structured representation of the machine learning models in their prospective submissions to the regulatory bodies.

Keywords: machine learning, artificial neural networks, quality by design, pharmaceutical development, process analytical technologies

Introduction

There is an ever-increasing need for the rapid development of pharmaceutical products, that can greatly rely on powerful computational methods. As in other research fields, artificial intelligence (AI), especially machine learning (ML) algorithms, has proven its great potential for deciphering complex relationships between multivariable data that are generated in the pharmaceutical development. Both formulation composition and processing parameters can be efficiently optimized, together with the minimized variability in the final products' quality. ML modeling also provides the ability to analyze unstructured datasets and predict formulation and process properties for any given combination of independent variables. These are, therefore, undoubtedly important tools that coupled with the conventional statistical and regression methods create a data science platform (Figure 1). In addition to the better understanding of the pharmaceutical formulation and its manufacturing technology, data science is also resource-efficient.

The introduction of process analytical technologies (PAT) has facilitated the acquisition of process-related data, that is high in volume, variety and velocity of generation. This is related to the *big data* concept (1), whereby ML tools can provide a greater understanding of the data generated during pharmaceutical processing, through the identification of sensitivities and interdependencies of variables. Thereby, the variability of final products (dosage forms), that could potentially affect therapeutic efficacy, can be reduced.
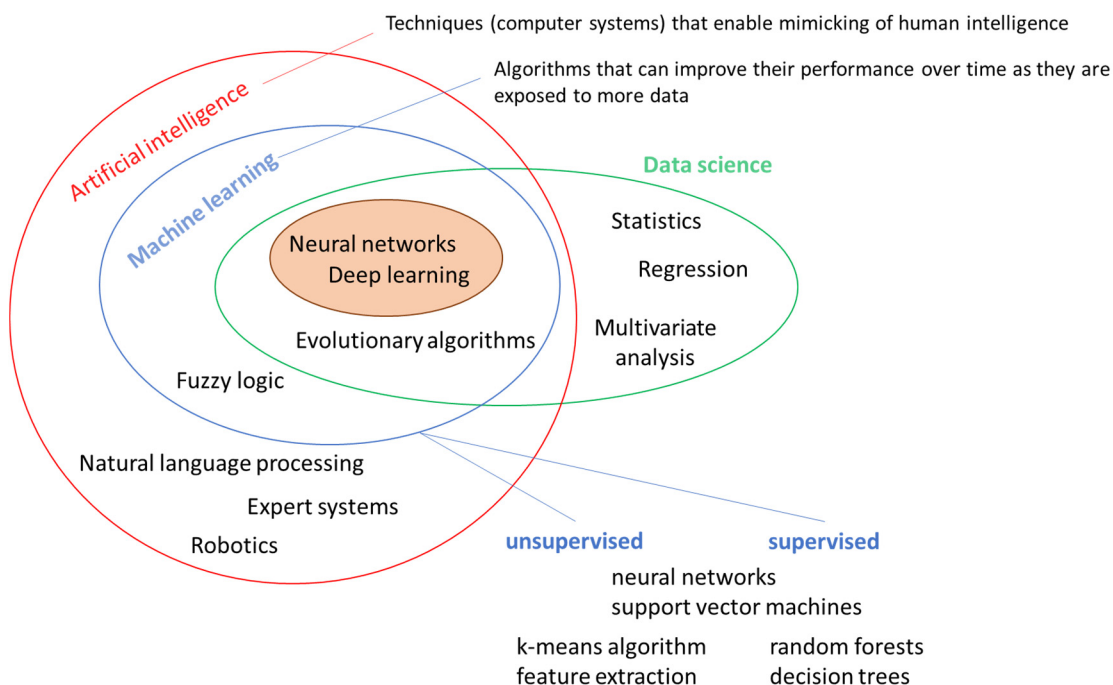

Figure 1.   Relationships between different methodologies in data mining
Slika 1.   Povezanost različitih tehnika za obradu podataka

Overview of the available AI methodologies

Concepts of artificial intelligence and machine learning have been introduced in the middle of the 20th century. The versatility of their application in various fields, including healthcare, development of new medicines, and (bio)medicine in general, has been growing ever since (2). ML can be based on supervised and unsupervised learning algorithms. In the case of supervised learning, data are designated as input (independent) and output (dependent), and the algorithm searches to find the best relationship that can be used for generalization and predictions. These methods are comparable to the conventional regression techniques. Unsupervised learning, on the other hand, is based on the assessment of the dataset as a whole and recognition of patterns and features, that provide the means for further clustering, reduction of dimensionality, etc. The type of the analyzed data (e.g. discrete or continuous, binary or multiple classes), as well as the type of the model (e.g. parametric or non-parametric), are also relevant for the selection of the appropriate machine learning algorithm. Artificial neural networks (ANNs), as the most frequently used machine learning algorithms, can be used both for regression and classification. Damiati has provided a comprehensive comparison of different machine learning methods that are used in pharmaceutical research (1).

There are many different types of ANNs. Multi-Layered Perceptron (MLP) is the most often used, as one of the simplest yet powerful networks. It is schematically represented in Figure 2. Input data are fed into the network through the input layer neurons whereby the number of neurons represents the number of input variables. Input data are further transformed and analyzed via activation functions in the hidden layer. As represented in Figure 2, sigmoid or hyperbolic functions are often used, but many others can be also implemented (3). This versatility of functions that can be applied to model the data is one of the most important reasons why ANNs can outperform conventional regression methods. Non-linear equations that are used by ANNs are fitted through a number of iterations, in order to capture the variability within the data. This is referred to as the training process, and there is a number of methods that can be used to adjust the synapses, i.e. networks´ weights and coefficients. Backpropagation algorithm, for instance, calculates the difference between the actual and predicted output, and based on the obtained error adjusts the weights and coefficients. The process is repeated until the network converges to the optimum solution. Depending on the type of connections networks can be fully or partially connected. Furthermore, the type of activation functions may vary in different layers. ANNs can have more than one hidden layer, and also the recurrence of the signal may occur (3).

Deep learning is a more recent concept, based on complex ANN structures. In essence, deep learning neural networks have a large number of hidden layers (at least three but usually more), each layer has many nodes, and these networks are predominately built for large datasets. Deep learning networks include convolutional, recurrent, and fully connected networks and are discussed in more detail elsewhere (4).
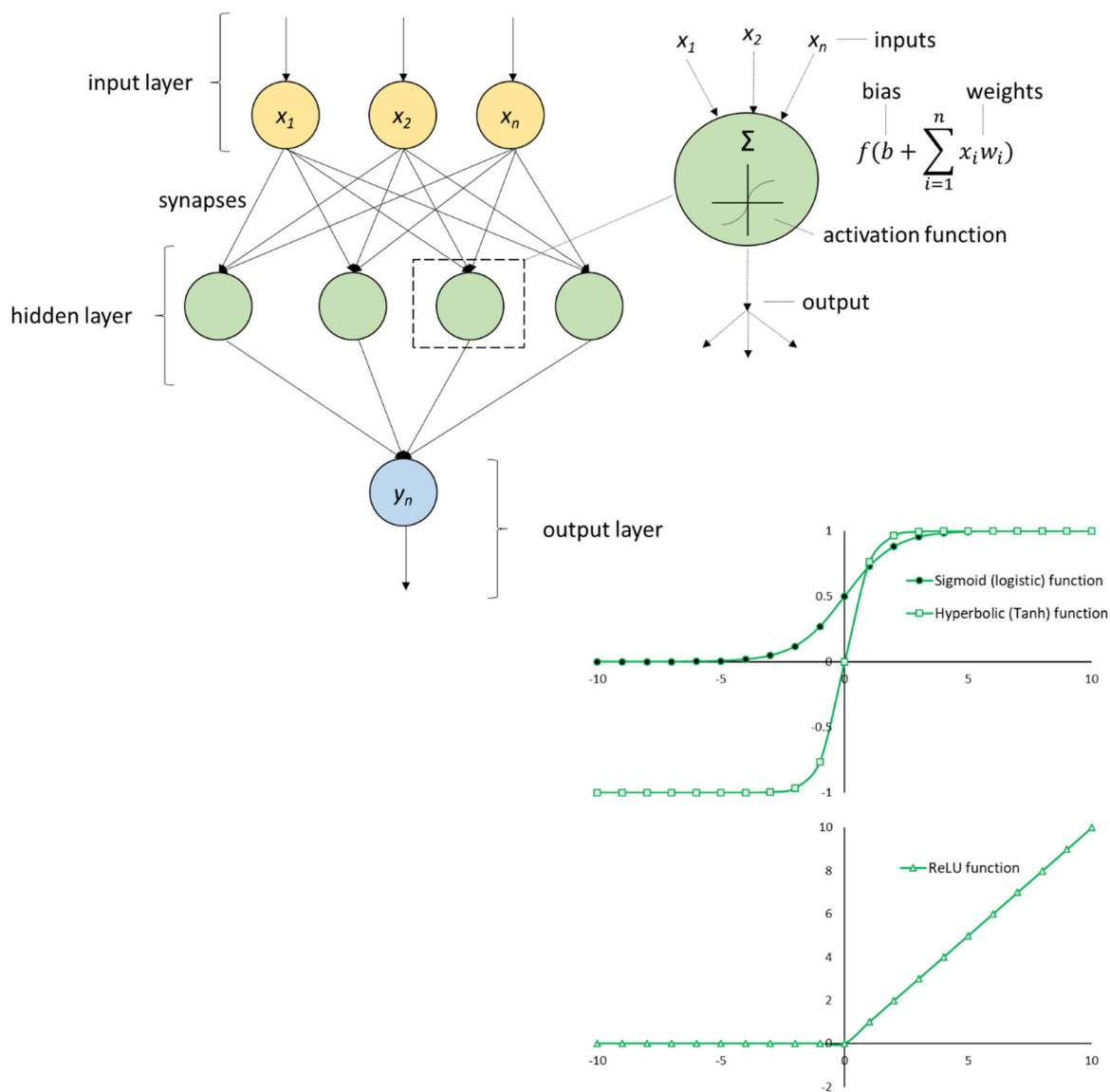
Figure 2. Representation of an artificial neural network with the most commonly used activation functions (adapted from (3))

Slika 2. Prikaz strukture neuronske mreže i najčešće korišćene aktivacione funkcije (prilagođeno iz (3))

Other ML algorithms, such as neuro-fuzzy logic, symbolic regression, decision trees, genetic programming, etc., have also attracted attention. Fuzzy models provide the opportunity to model relationships between the input and output variables with *if-then* logical statements (5). Prerequisites for such a modeling approach are membership functions and fuzzy sets. For more details on neuro-fuzzy logic and neuro-fuzzy inference systems, the interested reader is referred to (6) for more details. Genetic programming (GP) is based on the natural selection theory, whereby the GP algorithm generates

populations of fitting equations and then searches for the optimal population by applying genetic operations such as "reproduction", "crossover" or "mutation" (7).

Machine learning algorithms may be used individually or combined in structures denoted as forests or ensembles. Ensembles are extremely powerful in the case of multicriteria optimization for large datasets.

Illustrative examples of ML tools applied in the context of pharmaceutical formulation development and related technologies

The current approach to pharmaceutical development should be based on the quality by design principles (QbD) (8). The first step in the QbD approach is the definition of the quality target product profile (QTPP), followed by the identification of the critical quality attributes (CQA) of the product. The most important aspect of the QbD is the establishment of relationship(s) between the critical material attributes (CMA) and/or critical process parameters (CPP) that affect CQAs. Once these relationship(s) are identified and quantified, design space can be appointed, providing the opportunity to optimize and continuously control the product's quality. Historically, quantitative analysis and appointment of the design space were based on experimental design, regression methods, and conventional statistical analysis. However, there are actually no limits on methods that can be used for quantitative assessments in the QbD context. In fact, a whole array of techniques is available, under the data science umbrella (Figure 1), that can efficiently be used for a variety of QbD elements. ML algorithms, especially ANNs, have been used in numerous examples of QbD-based pharmaceutical development, specifically due to their non-linear nature and the ability to capture complex relationships between CMAs and/or CPPs with CQAs for various pharmaceutical dosage forms (9–17).

Simões et al. (18) have recently published a study on ANNs applied to QbD-based development of a poorly soluble and poorly permeable drug (class IV drug according to the BCS – Biopharmaceutics classification system) tablet formulation that was manufactured in industrial settings and compared to the reference product in bioequivalence studies. In short, after the initial risk assessment, the following CMAs and CPPs were identified as critical for the dissolution profile as a CQA: particle size distribution, tablets´ hardness, impeller speed, mesh size for sieving of the dried granules, granulation time and granulation liquid amount. Fully connected MLP networks were trained and validated. ANNs were then used to set product and process specifications, taking into account similarity factors for the predicted dissolution profiles. This allowed the establishment of the control strategy for the entire process. The optimal ANN model was validated on three industrial-scale batches manufactured with three different batches of the milled drug. The ANN model has successfully predicted dissolution profiles for the manufactured batches. This study represents a true power of ML algorithms in pharmaceutical products development. Lao et al. (19) have prepared an in-depth review of the application of ML in solid oral dosage form development in both academia and industry for the last three decades.

Table I represents selected examples of recently published reports on ANN models used in pharmaceutical development (for formulation and/or process optimization). Additional reviews of ANNs applied in pharmaceutical development are available (20–26).

Table I   Selected examples of ANN models used in pharmaceutical development (for formulation and/or process optimization)

Tabela I   Odabrani primeri modela zasnovanih na veštačkim neuronskim mrežama koji su korišćeni u razvoju lekova (za optimizaciju formulacije i/ili procesa)

ANN type(s)	Dataset properties	Brief comment on the network(s) purpose	Reference
kNN, single layer networks, RBF, SVM	73 excipients with specified seven physico-chemical properties	Different neural networks were compared in terms of their ability to identify excipients with the potential to be used as transdermal enhancers	(51)
MLP, SVM, random forest, extra trees and bagging	14 descriptors were used for several API datasets	Prediction of active pharmaceutical ingredients solubility in several solvents	(52)
MLP, RBF, Kohonen SOM	90 entries on formulation composition, compression load and seven tableting parameters	Understanding of tableting properties of lactose co-processed with lipid excipients	(53)
Tree based ensembles and symbolic regression methods	17 formulations, 2 compaction machines with 3 compaction pressures and 2 compaction speeds were used	Modeling of tablets tensile strength based on formulation composition and tableting process parameters	(54)
ANNs, SVM, random forests, decision trees, lightGBM, kNN, naïve Bayes classifiers and deep NNs	646 stability data of solid dispersions were collected and used for development of the classification models	Evaluation of the influence of formulations, process parameters and stability data, in order to predict the solid dispersions physical stability	(55)
SVM, backpropagation neural network, genetic algorithms, mind evolutionary algorithm	Four factors were evaluated (reflecting concentrations and type of materials) in a study according to the design of experiments	Assessment of the effect of the powder coating on the compactibility of the obtained materials	(56)
ANNs and deep neural networks	145 formulation data were collected that contained 23 active pharmaceutical ingredients	Prediction of the properties of directly compressed orally disintegrating tablets	(57)

Millen et al. (27) have compared multiple linear regression; stepwise, ridge and lasso regression; elastic network, regression trees and boosted regression trees as modeling techniques for the assessment of particle size distribution and tablets' quality attributes following wet granulation on different scales (from laboratory to full-scale production). Landin has demonstrated that the combination of neuro-fuzzy logic and genetic programing allows modeling of the wet granulation process for different sizes and geometries of the wet granulation equipment (28). Belič et al. have applied ANNs and fuzzy models to minimize the capping tendency by the optimization of the tableting process (5). ANNs were also compared to adaptive neuro-fuzzy inference system (ANFIS) and multiple linear regression, in terms of modeling the compaction performance of novel pharmaceutical excipients (thermally and chemically modified starches) (29). Gams et al. (30) have developed an ML based method, using decision trees and support vector machines (SVM), that evaluates material properties and processing parameters for the successful production of tablets.

Barmpalexis et al. (7) have compared multi-linear regression, particle swarm optimization (PSO) ANNs and genetic programming in the development of mini-tablets. PSO-ANNs were the only regression technique that was able to simultaneously model eight responses (describing powder and mini-tablets properties). PSO is an optimization tool based on bird flocking behavior (7).

Colombo has provided an in-depth overview of AI applications in sophisticated targeted drug delivery systems, also proposing complete product development roadmaps based on the QbD principles (31). Hassanzadeh et al. have presented the significance of AI in the design of drug delivery systems (25). Adir et al. (32) have provided a review on how AI can be coupled with nanotechnology to enable a more efficient development of the precision cancer medicine. Similarly, Egorov et al. (33) have described the role of AI, coupled with robotics, microfluidics and nanotechnology, in the design of liposomes and polymeric systems.

Lee et al. have presented an interesting approach to the patent literature review (34). Multilayer neural networks were developed in order to identify the trends of specific pharmaceutical technologies' emergingness. Similarly, Lin et al. have created a SVM-based classifier for the prediction of the licensing outcomes of pharmaceutical patents held by universities (35).

The superiority of deep learning over conventional machine learning approaches (including multiple linear regression, partial least square regression, support vector machine, artificial neural networks, random forest, and k-nearest neighbors) has been recently demonstrated (36). Yang et al. (36) have extracted experimental data for 131 oral fast disintegrating films and 145 sustained-release matrix tablet formulations. Disintegration times and dissolution profiles for the respective formulations have been accurately modelled and the appropriate generalization has been confirmed by external datasets. The analyzed dataset contained types and contents of active pharmaceutical ingredients and excipients, process parameters and *in vitro* properties of dosage forms.

Deep learning has been demonstrated comparable, after 10-fold cross-validation, to random forest, single tree and genetic algorithms, for prediction of drug release from poly-lactide-co-glycolide (PLGA) microspheres (37).

In addition to deep learning, other advanced machine learning algorithms have demonstrated their prediction efficiency. Light gradient boosting machine algorithm (lightGBM), as a high-performance boosting decision tree was used to predict complexation between cyclodextrins and active pharmaceutical ingredients (38), as well as complexion of berberine into phospholipid complexes (39). The same machine learning algorithm, lightGBM, was also used for the prediction of particle size and polydispersity index (PDI) of nanocrystals prepared by top-down methods (40). Apart from predictions of target variables, light gradient boosting machine also provided information on the relative contribution of the formulation factors and process parameters on the nanocrystals size and PDI (40). Decision-tree-based methods were demonstrated capable of predicting particle size of solid lipid nanoparticles as well (41).

LightGBM, coupled with natural language processing (NLP) and blockchain technology, was also used to develop a management and recommendation system for the drug supply chain (42).

Regulatory aspects of machine learning algorithms application in pharmaceutical development

There is a great interest and potential for the pharmaceutical industry to utilize machine learning algorithms in practically every aspect of the pharmaceutical products' lifecycle. Henstock has reviewed and recommended steps for the successful integration of artificial intelligence, in the general context, in the pharmaceutical industry (43).

Regarding the official guidelines and recommendations, Food and Drug Administration (FDA) has recently presented a discussion paper followed by the action plan published in 2021 (44), that is devoted to artificial intelligence and machine learning software used as a medical device. These documents provide valuable concepts, such as *Good Machine Learning Practice* and *Real-World Performance* that might pave the way for similar AI and ML applications in pharmaceutical development. According to the action plan (44), good machine learning practices refer to efficient data management, feature extraction, training, interpretability, evaluation and documentation. European Medicines Agency (EMA) has, on the other hand, initiated a *big data* task force, together with the Heads of Medicines Agencies (HMA). Presentations on the current state-of-the-art use of AI in medicines, from the perspective of the relevant stakeholders, are available on EMA's website (45). One of EMA's strategic goals is to strengthen the ability to validate AI algorithms, and to deal with big data in general, since it is inevitable that these tools are going to be increasingly applied by the pharmaceutical industry (46).

Danish Medicines Agency has published a proposal for the criteria and questions that can be expected to be asked for AI/ML algorithms across the various GxP-regulated

areas (47). It is stated that the proposal applies to static AI/ML algorithms that implement critical GxP-related functions and are trained using supervised learning.

Future expectations

Serialization (track and trace) could also rely on ML tools. Although it is not its primary purpose, the data generated by serialization could also be used in product development and for tracking patients' adherence. Also, ML algorithms could be used for the identification of falsified medicines (48).

More efficient manufacturing in the pharmaceutical industry can be expected with the integration of ML and PAT tools (49). Several such examples have been described in Table II. One of the greatest challenges related to successful PAT implementation is the analysis of high volume, multivariate data. Moreover, fast computations and decision making are of the utmost importance. These issues could be, potentially, solved by different ML tools. For example, Wong et al. (50) have developed a method based on recurrent neural networks that provides efficient regulation of critical quality attributes.

With many opportunities for the application of AI tools also come some obstacles and challenges, especially if algorithms and models are meant to be used in the mass production of medicines, either for pharmaceutical development or production process monitoring, or both. The challenges are related to the volume of data and speed of its accumulation; size of datasets; training/learning time, over or under-fitting of models, etc. (1). With the evolution of the big data concept and advances in computing capability, it is to be expected that the technical challenges will be reduced, but the necessity for critical considerations of AI-based models will still remain, as for any other modeling approach.

Table II Selected examples of machine learning (ML) algorithms used in conjunction with the process analytical technologies (PAT)

Tabela II Odabrani primeri algoritama mašinskog učenja koji su korišćeni uz alatke za procesnu analizu (procesne analitičke tehnologije)

Application domain	Specific purpose	Brief comment on the algorithm	Reference
Machine vision	On-line monitoring of crystal growth	Deep-learning based (convolutional neural network) image segmentation algorithm	(58)
	In-line monitoring of pellets' agglomeration during fluidized-bed coating	Clustering-based image segmentation and convolutional neural network for classification of detected pellets	(59)
	Image analysis for the control of the tablet coating quality	Support vector machine and convolutional neural networks were used for tablets' image classification into four clusters	(60)
Real-time prediction of CQAs	Prediction of the entire dissolution profile for modified release tablets based on NIR and Raman spectra of intact tablets	Feedforward, fully-connected neural networks were developed; datasets consisted of the NIR and Raman spectra of each of the intact tablet as inputs and the dissolution curves of the same tablets measured by the dissolution method as outputs	(61)
Process monitoring	In-line measurement of the air temperature distribution during a fluidized-bed granulation process	Feed-forward ANNs with the back-propagation algorithm were used to make predictions of the temperature distribution and the hydrodynamics of the bed during the granulation process	(62)
Biotechnology	Description of biomass and product formation rate based on several on-line process parameters	Parametric dynamic bioreactor model was integrated with neural networks of different architectures	(63)
	ANN model was developed for at-line prediction of a recombinant protein	Multilayer perceptron model was built in conjunction with the partial least squares model	(64)

List of abbreviations

AI – artificial intelligence

ANFIS – adaptive neuro-fuzzy inference system

ANN – artificial neural network

BCS – biopharmaceutics classification system

CMA – critical material attributes

CPP – critical process parameters

CQA – critical quality attributes

EMA – European Medicines Agency

FDA – Food and Drug Administration

GBM – gradient boosting machine algorithm

HMA – Heads of Medicines Agencies

kNN – k-nearest neighbor

ML – machine learning

MLP – multi-layered perceptron

NLP – natural language processing

PAT – process analytical technology

PDI – polydispersity index

PSO – particle swarm optimization

QbD – quality by design

QTPP – quality target product profile

RBF – radial basis function

SOM – self-organizing map

SVM – support vector machine

References

1. Damiati SA. Digital Pharmaceutical Sciences. AAPS PharmSciTech. 2020;21(6):206.
2. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Bohr A, Memarzadeh K, editors. Artificial Intelligence in Healthcare: Elsevier; 2020; p. 25-60.
3. Djuris J, Vidovic B, Ibric S. Release modeling of nanoencapsulated food ingredients by artificial intelligence algorithms. In: Jafari SM, editor. Release and Bioavailability of Nanoencapsulated Food Ingredients: Elsevier; 2020; p. 311-47.
4. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241-50.

5. 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.

6. Walia N, Singh H, Sharma A. ANFIS: Adaptive Neuro-Fuzzy Inference System- A Survey. Int J Comp App. 2015;123(13):32-8.

7. 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(1–2):166-76.

8. ICH Q8(R2). Pharmaceutical Development. [Internet]. Available from: https://database.ich.org/sites/default/files/Q8_R2_Guideline.pdf. [cited 2021 May 15].

9. Amasya G, Aksu B, Badilli U, Onay-Besikci A, Tarimci N. QbD guided early pharmaceutical development study: Production of lipid nanoparticles by high pressure homogenization for skin cancer treatment. Int J Pharm. 2019;563:110-21.

10. Barmpalexis P, Kachrimanis K, Georgarakis E. Solid dispersions in the development of a nimodipine floating tablet formulation and optimization by artificial neural networks and genetic programming. Eur J Pharm Biopharm. 2011;77(1):122-31.

11. Koletti AE, Tsarouchi E, Kapourani A, Kontogiannopoulos KN, Assimopoulou AN, Barmpalexis P. Gelatin nanoparticles for NSAID systemic administration: Quality by design and artificial neural networks implementation. Int J Pharm. 2020;578:119118.

12. Chansanroj K, Petrović J, Ibrić S, Betz G. Drug release control and system understanding of sucrose esters matrix tablets by artificial neural networks. Eur J Pharm Sci. 2011;44(3):321-31.

13. Miletić T, Ibrić S, Đurić Z. Combined Application of Experimental Design and Artificial Neural Networks in Modeling and Characterization of Spray Drying Drug: Cyclodextrin Complexes. Drying Tech. 2014;32(2):167-79.

14. Aksu B, Paradkar A, de Matas M, Özer Ö, Güneri T, York P. A quality by design approach using artificial intelligence techniques to control the critical quality attributes of ramipril tablets manufactured by wet granulation. Pharm Dev Technol. 2013;18(1):236-45.

15. Nguyen CN, Tran BN, Do TT, Nguyen H, Nguyen TN. D-Optimal Optimization and Data-Analysis Comparison Between a DoE Software and Artificial Neural Networks of a Chitosan Coating Process onto PLGA Nanoparticles for Lung and Cervical Cancer Treatment. J Pharm Innov. 2019;14(3):206-20.

16. Onuki Y, Kawai S, Arai H, Maeda J, Takagaki K, Takayama K. Contribution of the Physicochemical Properties of Active Pharmaceutical Ingredients to Tablet Properties Identified by Ensemble Artificial Neural Networks and Kohonen's Self-Organizing Maps. J Pharm Sci. 2012;101(7):2372-81.

17. Kinnunen H, Hebbink G, Peters H, Shur J, Price R. Defining the Critical Material Attributes of Lactose Monohydrate in Carrier Based Dry Powder Inhaler Formulations Using Artificial Neural Networks. AAPS PharmSciTech. 2014;15(4):1009-20.

18. Simões MF, Silva G, Pinto AC, Fonseca M, Silva NE, Pinto RMA, et al. Artificial neural networks applied to quality-by-design: From formulation development to clinical outcome. Eur J Pharm Biopharm. 2020;152:282-95.

19. Lou H, Lian B, Hageman MJ. Applications of Machine Learning in Solid Oral Dosage Form Development. J Pharm Sci. 2021; doi: 10.1016/j.xphs.2021.04.013.

20. Rowe RC, Roberts RJ. Artificial intelligence in pharmaceutical product formulation: neural computing and emerging technologies. Pharm Sci Technol Today. 1998;1(5):200-5.

21. Shao Q, Rowe RC, York P. Investigation of an artificial intelligence technology—Model trees. Eur J Pharm Sci. 2007;31(2):137-44.

22. Colbourn EA, Rowe RC. Novel approaches to neural and evolutionary computing in pharmaceutical formulation: challenges and new possibilities. Future Med Chem. 2009;1(4):713-26.

23. 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.

24. Djuris J, Ibric S, Djuric Z. Neural computing in pharmaceutical products and process development. In: Djuris J, editor. Computer-Aided Applications in Pharmaceutical Technology: Elsevier; 2013; p. 91-175.

25. Hassanzadeh P, Atyabi F, Dinarvand R. The significance of artificial intelligence in drug delivery system design. Adv Drug Deliv Rev. 2019;151–152:169-90.

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

27. Millen N, Kovačević A, Djuriš J, Ibrić S. Machine Learning Modeling of Wet Granulation Scale-up Using Particle Size Distribution Characterization Parameters. J Pharm Innov. 2020;15(4):535-46.

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

29. 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.

30. Gams M, Horvat M, Ožek M, Luštrek M, Gradišek A. Integrating Artificial and Human Intelligence into Tablet Production Process. AAPS PharmSciTech. 2014;15(6).

31. Colombo S. Applications of artificial intelligence in drug delivery and pharmaceutical development. In: Bohr A, Memarzadeh K, editors. Artificial Intelligence in Healthcare: Elsevier; 2020; p. 85-116.

32. Adir O, Poley M, Chen G, Froim S, Krinsky N, Shklover J, et al. Integrating Artificial Intelligence and Nanotechnology for Precision Cancer Medicine. Adv Mat. 2020;32(13):1901989.

33. Egorov E, Pieters C, Korach-Rechtman H, Shklover J, Schroeder A. Robotics, microfluidics, nanotechnology and AI in the synthesis and evaluation of liposomes and polymeric drug delivery systems. Drug Deliv Trans Res. 2021;11(2):345-52.

34. Lee C, Kwon O, Kim M, Kwon D. Early identification of emerging technologies: A machine learning approach using multiple patent indicators. Technol Forecast Soc Change. 2018;127:291-303.

35. Lin H-H, Ouyang D, Hu Y. Intelligent Classifier: a Tool to Impel Drug Technology Transfer from Academia to Industry. J Pharm Innov. 2019;14(1):28-34.

36. Yang Y, Ye Z, Su Y, Zhao Q, Li X, Ouyang D. Deep learning for in vitro prediction of pharmaceutical formulations. Acta Pharm Sin B. 2019;9(1):177-85.

37. Ekins S. The Next Era: Deep Learning in Pharmaceutical Research. Pharm Res. 2016;33(11):2594-603.

38. Zhao Q, Ye Z, Su Y, Ouyang D. Predicting complexation performance between cyclodextrins and guest molecules by integrated machine learning and molecular modeling techniques. Acta Pharm Sin B. 2019;9(6):1241-52.

39. Gao H, Ye Z, Dong J, Gao H, Yu H, Li H, et al. Predicting drug/phospholipid complexation by the lightGBM method. Chem Phys Lett. 2020;747:137354.

40. He Y, Ye Z, Liu X, Wei Z, Qiu F, Li H-F, et al. Can machine learning predict drug nanocrystals? J Control Release. 2020;322:274-85.

41. Öztürk AA, Gündüz AB, Ozisik O. Supervised Machine Learning Algorithms for Evaluation of Solid Lipid Nanoparticles and Particle Size. Comb Chem High Throughput Screen. 2018;21(9):693-9.

42. Abbas K, Afaq M, Ahmed Khan T, Song W-C. A Blockchain and Machine Learning-Based Drug Supply Chain Management and Recommendation System for Smart Pharmaceutical Industry. Electronics. 2020;9(5):852.

43. Henstock PV. Artificial Intelligence for Pharma: Time for Internal Investment. Trends Pharmacol Sci. 2019;40(8):543-6.

44. Artificial Intelligence and Machine Learning in Software as a Medical Device [Internet]. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device. [cited 2021 May 15]

45. Joint HMA/EMA workshop on artificial intelligence in medicines regulation. [Internet]. Available from:https://www.ema.europa.eu/en/events/joint-hmaema-workshop-artificial-intelligence-medicines-regulation [cited 2021 May 15].

46. EMA Regulatory Science to 2025 Strategic reflection. [Internet]. Available from: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf [cited 2021 May 15].

47. Suggested criteria for using AI/ML algorithms in GxP. [Internet]. Available from: https://laegemiddelstyrelsen.dk/en/licensing/supervision-and-inspection/inspection-of-authorised-pharmaceutical-companies/using-aiml-algorithms-in-gxp/ [cited 2021 May 15].

48. Rasheed H, Höllein L, Holzgrabe U. Future Information Technology Tools for Fighting Substandard and Falsified Medicines in Low- and Middle-Income Countries. Front Pharmacol. 2018;9:995.

49. Klemenčič J, Mihelič J. Application of Algorithms and Machine Learning Methods in Pharmaceutical Manufacture. IPSI Trans Internet Res. 2019;15(1):16–22.

50. Wong W, Chee E, Li J, Wang X. Recurrent Neural Network-Based Model Predictive Control for Continuous Pharmaceutical Manufacturing. Mathematics. 2018;6(11):242.

51. Moss GP, Shah AJ, Adams RG, Davey N, Wilkinson SC, Pugh WJ, et al. The application of discriminant analysis and Machine Learning methods as tools to identify and classify compounds with potential as transdermal enhancers. Eur J Pharm Sci. 2012;45(1–2):116-27.

52. Boobier S, Hose DRJ, Blacker AJ, Nguyen BN. Machine learning with physicochemical relationships: solubility prediction in organic solvents and water. Nat Commun. 2020;11(1):5753.

53. 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.

54. Khalid MH, Tuszynski PK, Szlek J, Jachowicz R, Mendyk A. From Black-Box to Transparent Computational Intelligence Models: A Pharmaceutical Case Study. In: 2015 13th International Conference on Frontiers of Information Technology (FIT). IEEE; 2015.

55. Han R, Xiong H, Ye Z, Yang Y, Huang T, Jing Q, et al. Predicting physical stability of solid dispersions by machine learning techniques. J Control Release. 2019;(311–2):16-25.

56. Lou H, Chung JI, Kiang Y-H, 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.

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

58. Chen S, Liu T, Xu D, Huo Y, Yang Y. Image based Measurement of Population Growth Rate for L-Glutamic Acid Crystallization. In: 2019 Chinese Control Conference (CCC). IEEE; 2019.

59. Mehle A, Likar B, Tomaževič D. In-line recognition of agglomerated pharmaceutical pellets with density-based clustering and convolutional neural network. IPSJ Trans Comput Vis Appl. 2017;9(1):1-6.

60. Hirschberg C, Edinger M, Holmfred E, Rantanen J, Boetker J. Image-Based Artificial Intelligence Methods for Product Control of Tablet Coating Quality. Pharmaceutics. 2020;12(9):877.

61. Nagy B, Petra D, Galata DL, Démuth B, Borbás E, Marosi G, et al. Application of artificial neural networks for Process Analytical Technology-based dissolution testing. Int J Pharm. 2019;567:118464.

62. 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;88:219-32.

63. von Stosch M, Hamelink J-M, Oliveira R. Hybrid modeling as a QbD/PAT tool in process development: an industrial E. coli case study. Bioprocess Biosyst Eng. 2016;39(5):773-84.

64. Chiappini FA, Teglia CM, Forno ÁG, Goicoechea HC. Modelling of bioprocess non-linear fluorescence data for at-line prediction of etanercept based on artificial neural networks optimized by response surface methodology. Talanta. 2020;210:120664.

Pregled primene algoritama mašinskog učenja u farmaceutskoj tehnologiji

Jelena Djuriš, Ivana Kurćubić, Svetlana Ibrić

Univerzitet u Beogradu – Farmaceutski fakultet, Katedra za farmaceutsku tehnologiju i kozmetologiju, Vojvode Stepe 450, 11221 Beograd, Srbija

Autor za korespondenciju: Jelena Djuriš, e-mail: jelena.djuris@pharmacy.bg.ac.rs

Kratak sadržaj

Algoritmi mašinskog učenja, kao i veštačka inteligencija u širem smislu, su veoma značajni i primenjuju se u razne svrhe u okviru farmaceutske tehnologije. Počevši od razvoja formulacija, preko izuzetnog potencijala za integraciju u koncept dizajna kvaliteta (engl. *Quality by design*), algoritmi mašinskog učenja omogućavaju bolje razumevanje uticaja kako formulacionih faktora tako i odgovarajućih procesnih parametara. Algoritmi mašinskog učenja mogu biti od naročitog značaja i za analizu velikog obima podataka koji se generišu korišćenjem procesnih analitičkih tehnologija. U ovom radu su ukratko predstavljene veštačke neuronske mreže, kao jedan od najčešće korišćenih algoritama mašinskog učenja. Prikazani su procesi treninga i testiranja mreža, kao i ilustrativni primeri algoritama primenjenih za različite potrebe razvoja i/ili optimizacije farmaceutskih formulacija i postupaka njihove izrade. Takođe, dat je i pregled budućih trendova u ovoj oblasti, kao i novijih studija o sofisticiranim metodama, poput dubokih neuronskih mreža, i *light gradient boosting* algoritma. Zainteresovani čitaoci se takođe upućuju na nekoliko zvaničnih dokumenata (vodiča), po uzoru na koje mogu da se očekuju i preporuke za strukturiranu prezentaciju modela mašinskog učenja koji će se podnositi regulatornim telima u okviru dokumentacije koja se priprema za potrebe registracije novih lekova.

Ključne reči: mašinsko učenje, veštačke neuronske mreže, razvoj lekova, dizajn kvaliteta, procesne analitičke tehnologije