Original Research Paper

Predicting oral disintegrating tablet formulations by neural network techniques ☆

Check for updates

Run Han[a], Yilong Yang[a,b], Xiaoshan Li[b], Defang Ouyang[a,*]

[a] *State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences (ICMS), University of Macau, Macau 999078, China*
[b] *Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau 999078, China*

ARTICLE INFO

ABSTRACT

Oral disintegrating tablets (ODTs) are a novel dosage form that can be dissolved on the tongue within 3 min or less especially for geriatric and pediatric patients. Current ODT formulation studies usually rely on the personal experience of pharmaceutical experts and trial-and-error in the laboratory, which is inefficient and time-consuming. The aim of current research was to establish the prediction model of ODT formulations with direct compression process by artificial neural network (ANN) and deep neural network (DNN) techniques. 145 formulation data were extracted from Web of Science. All datasets were divided into three parts: training set (105 data), validation set (20) and testing set (20). ANN and DNN were compared for the prediction of the disintegrating time. The accuracy of the ANN model have reached 85.60%, 80.00% and 75.00% on the training set, validation set and testing set respectively, whereas that of the DNN model were 85.60%, 85.00% and 80.00%, respectively. Compared with the ANN, DNN showed the better prediction for ODT formulations. It is the first time that deep neural network with the improved dataset selection algorithm is applied to formulation prediction on small data. The proposed predictive approach could evaluate the critical parameters about quality control of formulation, and guide research and process development. The implementation of this prediction model could effectively reduce drug product development timeline and material usage, and proactively facilitate the development of a robust drug product.

1.   Introduction

Oral dosage forms are always the most widely used dosage form because of their convenience of self-administration, good stability, accurate dosing and easy manufacturing [1]. However, swallowing difficulty of the pediatric or geriatric patient is a big concern for conventional tablets. Dysphagia is observed in about 35% of the general population among all age groups, as well as in up to 40% of the elder population and 18%–22% of all patients in long-term care facilities [2]. To overcome the difficulty in swallowing, oral disintegrating tablets (ODTs) have been developed since the 1990s [3,4]. ODTs are designed to be dissolved on the tongue rather than swallowed whole as conventional tablets [5,6]. The disintegrating time of ODTs is within 3 min or less in the saliva without the intake of water [7,8]. In recent years, there is the growing demand about good ODT formulations with new disintegrants and convenient preparation methods. There are three major techniques which are widely used for ODT manufacturing: freeze drying, tablet molding, and tablet compression [9,10]. Comparing with many other preparation methods, direct compression is most widely used because of its most effective and simplest process [11]. The formulations of ODTs with direct compression method usually contain the filler, binder, disintegrant, lubricant and solubilizer [12]. Therefore, formulation design of ODTs is critical to minimize the disintegrating time with good tablet quality.

Current pharmaceutical formulation development usually depends on experimental trial-and-error by personal experiences of formulation scientists, which is inefficient and time-consuming. To improve the efficiency of formulation screening, the SeDeM diagram expert system was developed to optimize formulations [13]. SeDeM diagram expert system was able to evaluate the influence of every excipient on the final formulation for direct compression based on the experimental study and quantitative characterization parameters [14]. Then this expert system considered the type of excipients and physicochemical properties to output a recommended formu-

lation. Moreover, the mathematical analysis of SeDeM was able to recommend not only formulation components but also the optimal ratios of excipients [14,15]. Firstly, 43 excipients were investigated the suitability for direct compression, especially the compressibility of disintegrants. According to the ICHQ8, the suitability was described as these parameters: bulk density, tapped density, inter-particle porosity, Carr index, cohesion index, Hausner ratio, angle of repose, powder flow, loss on drying, hygroscopicity, particle size and homogeneity index. The SeDeM system could show the profile of every excipient and evaluate how suitable it can be used for direction compression [12]. According to the predicted result and combining with the experimental study, 8 excipients with the better properties were chosen to make a comparison using the new expert system. Compared with the old system, the new system could quantify the compressibility index of every excipient with the higher precision [16]. For example, ibuprofen ODT formulations were investigated with the suitability of 21 excipients and obtained the final SeDeM diagram with 12 parameters [17]. Current SeDeM method just focused on the recommended formulation, but it cannot quantitatively predict the disintegrating time of ODT formulations. With the challenge of pharmaceutical research, we need to establish a prediction method to assist experts evaluate the performance of ODT formulations.

The neural network is a wonderful biologically-inspired model that learns from observational data. That is an artificial network with seriously connected units by simulating the neural structure of the brain [18]. Neural network has been applied to solve problems in many fields, such as voice recognition and computer vision. Artificial neural network and deep neural network are two widely used neural networks, as shown in Figs. 1 and 2 [19]. ANN is a simple neuron network with only one hidden layer, while DNN is a more powerful technique with many complex layers to reach the high-level data representation. In pharmacology and bioinformatics research, ANN also has been used over two decades, included prediction of protein secondary structure and quantitative structure–activity relationship [20]. As the pharma-
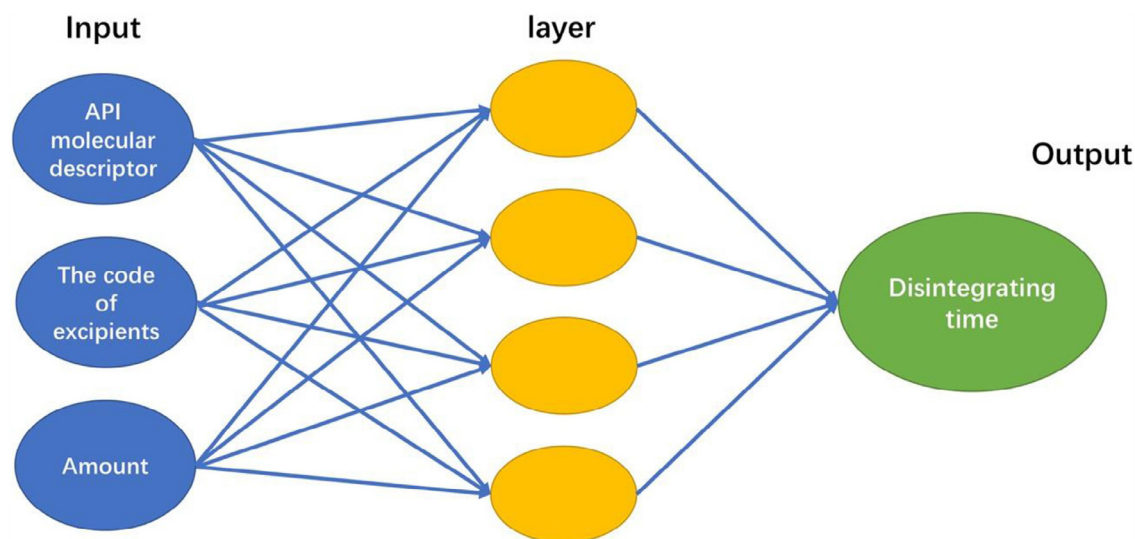

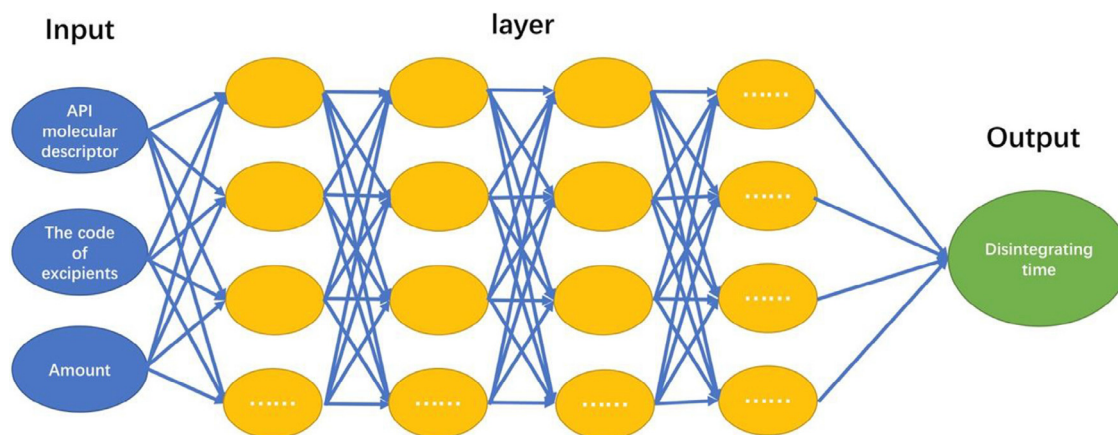**Fig. 1 – The network structure of ANN.**

Fig. 2 – The network structure of DNN.

ceutical research, the prediction models were developed for break force and disintegration of tablet formulation by ANN, genetic algorithm, support vector machine and random forest approaches [21]. Another ANN example was quantitative structure activity relationships (QSAR) of antibacterial activity study [22,23]. DNN is a type of representation learning with multiple levels of neural networks. Unlike the traditional ANN with manual feature extraction, deep-learning can automatically extract feature and even transform low-level representation to more abstract level without any feature extractor [24]. Moreover, deep-learning is more sensitive to irrelevant and particular minute variations with complicated parameters of the network, which could reach higher accuracy rather than the conventional machine learning algorithms [19]. In recent years, DNN has been applied in pharmacy research, such as drug design, drug-induced liver injury and virtual screening [25]. In most cases, deep-learning could generate a novel and complex system to represent various objects through molecular descriptor so that it would be very helpful for drug discovery and prediction [26]. Junshui Ma et al. extracted data from internal Merck data and included on-target and absorption, distribution, metabolism, excretion (ADME), each molecular was described as serious features. Finally, they used deep neural nets to evaluate QSAR and the result was better than random forest commonly used [27].

The aim of current research was to establish the quantitative prediction model of the disintegrating time of ODT formulations with direct compression process by ANN or DNN.

2. Methodology

2.1. Data extraction

Formulation data collection was the foundation of building the prediction model. To ensure the data reliability, the keyword search strategy was used in Web of Science database. The synonym strings of keywords were used, such as "oral" + "disintegrating" + "tablets" with 461 results, "fast" + "disintegrating" + "tablets" with 407 results, "rapidly" + "disintegrating" + "tablets" with 266 results, and "oral" + "dispersible" +

"tablets" with 84, respectively. Among these results, only research articles were selected for further data extraction. After the manual screening, 145 direct compressed ODT formulations with the disintegrating time were extracted including 23 active pharmacological ingredients (APIs) groups for our model, as shown in Table S1. All APIs were described as nine molecular parameters, including molecular weight, XLogP3, hydrogen bond donor count, hydrogen bond acceptor count, rotatable bond count, topological polar surface area, heavy atom count, complexity and logS. According to the function of excipients, all excipients were divided into five categories: filler, binder, disintegrant, lubricant, and solubilizer. Each type of excipient was individually coded for further training. The formulation data included API molecular descriptors and its amount, the type of encoded excipients and its amount, manufacture parameters (e.g. the hardness, friability, thickness and tablet diameter) and the disintegrating time of each formulation.

2.2. Dataset classification: Training set, validation set and testing set

To ensure good prediction ability of computational model, especially in the small amount of pharmaceutical data, the dataset should be carefully divided into three parts, including training set, validation set and testing set. The three datasets strategy is an effective way to test the accuracy on new data out of our datasets. In details, the training set is for training model and the validation set is used for adjusting the parameters and finding the best model, while testing set shows the prediction accuracy on real unknown data from the datasets, as shown in Fig. 3 . Therefore, how to select data for three datasets appropriately is the key step. Compared with random selection, manual selection and maximum dissimilarity algorithm selection, the improved maximum dissimilarity algorithm (MD-FIS) is the best choice. MD-FIS is based on the maximum dissimilarity algorithm considering small group data in the whole dataset, it will avoid selecting data mostly from small group and ensure the representation of validation and test set.
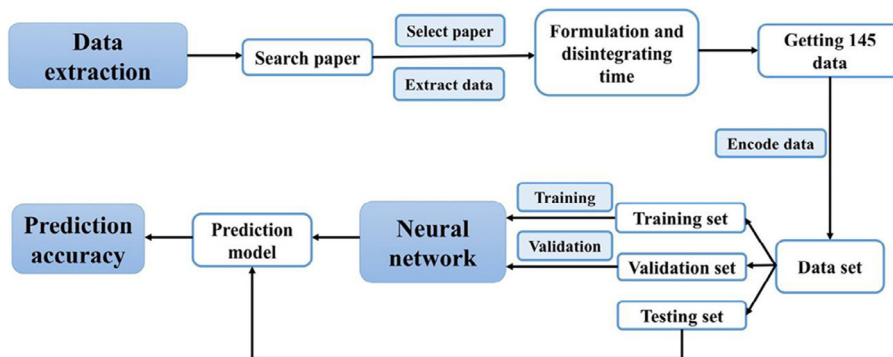
Fig. 3 – The flowchart of establishing model.

2.3. Hyperparameters of artificial neutral network and deep neural network

The prediction model for ODTs was trained by ANN and DNN, respectively. In the training process, all data are normalized and then divided into three sets with our previous proposed *MD-FIS* selection algorithm in R language. For ANN and DNN networks, Deeplearning4j machine learning framework (https://deeplearning4j.org/) was used to train prediction models. All the source codes can be found on the website (http://ml.mydreamy.net/pharmaceutics/ODT.html). The ANN model in Fig. 1 with termination condition at 15,000 epochs and hidden nodes is 200. The deep-learning process in Fig. 2 uses full-connected deep feedforward networks including ten layers with 2000 epochs. This neural network contains 50 hidden nodes on each layer. All networks choose *tanh* as the activation function except the last layer with *sigmoid* activation function. Learning rate is set to 0.01. Batch gradient descent with the 0.8 *momentum* is used for training the networks.

Note that epoch indicates how many times the dataset is used for training. Feedforward network means that the output of the network is computed layer-by-layer from one-direction without any inside loop. Learning rate impacts how fast the network will be convergent. Batch gradient descent is a training strategy to use all datasets to train the model at each time. Momentum indicates how much the speed will be kept in each training step.

2.4. Pharmaceutical evaluation criterion

European Pharmacopeia defined that ODT could disintegrate within 3 min in the mouth before being swallowed. In all our formulation data, the disintegrating time ranges from 0 s to 100 s. Usually, the successful prediction in pharmaceutics is that absolute error is less than 10%. Thus, a good model is that the prediction deviation of the disintegrating time is not more than 10 s. The accuracy of prediction disintegrating time is the percentage of successful prediction to total predictions:

$$Accuracy_{PDT} = \frac{Number(|f' - f| \le 10)}{A} ll\ Predictions$$

where, f' is the prediction value, f is the label (real) value. All predictions are the number of predicted data.

3. Results and discussion

Fig. 4 showed the label (true) value and predictive value of disintegrating time on ANN model (A. training set; B. validation set; C. testing set), while indicated the true value and predictive value of disintegrating time on DNN model (D. training set; E. validation set; F. testing set). As shown in Fig. 4, the training set and validation set of both ANN and DNN showed good results. As Table 1 shows, the predictive accuracy of ANN model is 85.60% on training set and 80.00% on validation set, while the DNN model is 85.60% and 85.00%, respectively. However, the testing set of ANN with only 75.00% accuracy is lower than that of DDN (80.00%), which indicated that DNN is able to sign"antly better predict real unknown data than ANN.

As the result shows, ANN is an efficient network for training prediction model within the adjustment of validation set, reaching a high accuracy on training set and validation set. However, when predicting real unknown data, the accuracy of testing set dropped significantly, which is called overfitting in machine learning. DNN performs well in all three datasets with over 80% accuracy and predicted stably with average value, which is more capable of establishing a better prediction model for ODT than ANN.

When analyzing the different network structures between ANN and DNN, ANN just includes one hidden layer, while DNN includes ten layers with 2000 epochs and each layer contains 50 hidden nodes. Thus, DNN could extract the feature of data with higher level and give a more accurate predictive result. It is unsurprising that DNN, as an innovative and effective technique for pharmaceutical research, can provide a higher accuracy prediction about disintegrating time than ANN. Thus, the desired DNN with the proposed MD-FIS selection algorithm can be used to achieve good predictive results on pharmaceutical formulations with small data.

In order to ensure a satisfied prediction accuracy, two key factors are to be considered: data and algorithm. The first issue is the reliable data in pharmaceutical research. Deeplearning attempts to learn these characteristics to make better representations and create models from reliable data. Thus, data extraction is a critical step. In current research, reliable formulation datasets were manually extracted and labeled from the research articles of Web of Science by experienced pharmaceutical scientists.
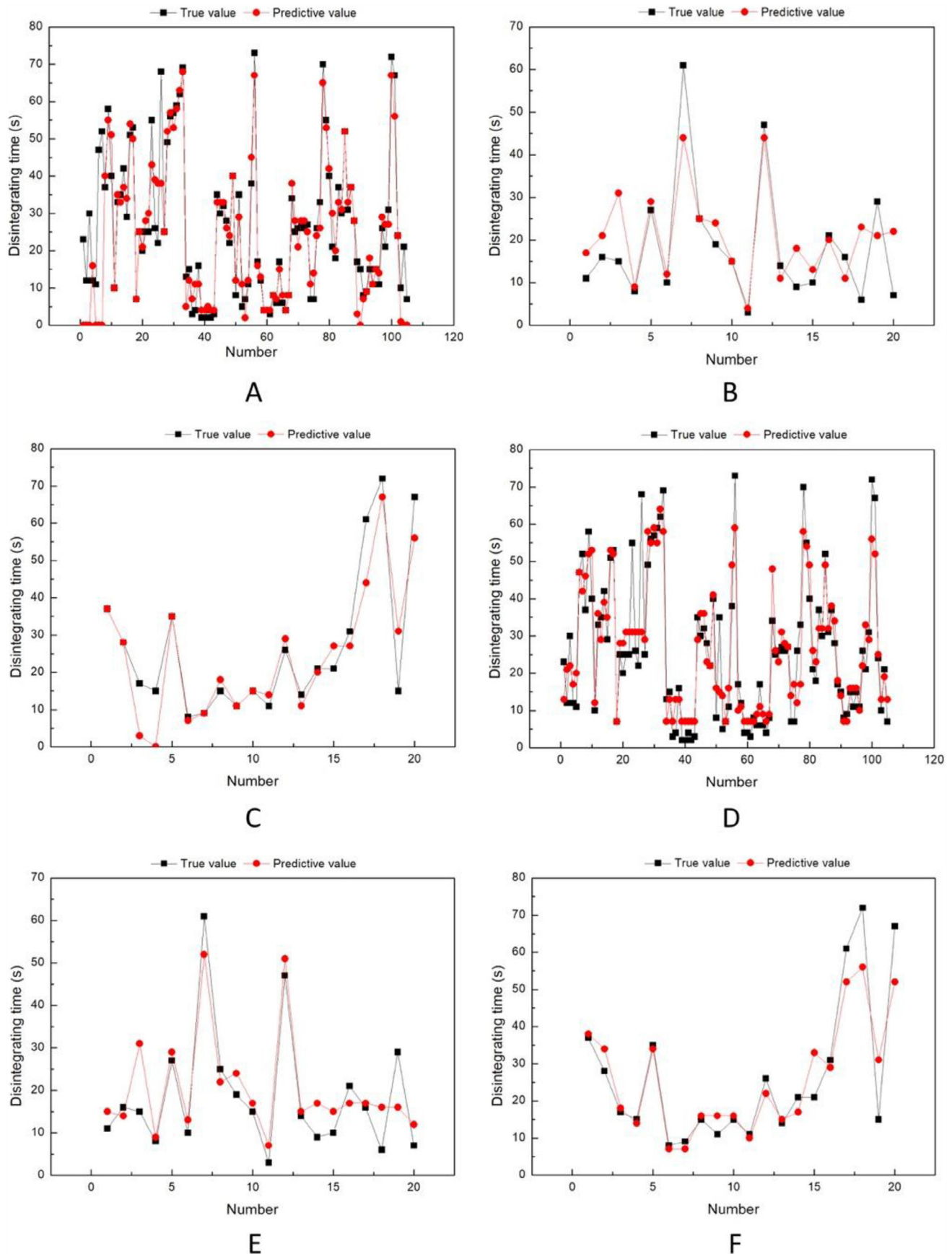
Fig. 4 – The true value and predictive value on dataset: (A) the true value and predictive value of training set and (B) validation set and (C) testing set on ANN model. (D) the true value and predictive value of training set and (E) validation set and (F) testing set on DNN model.

Table 1 – The accuracies of OFDT on training, validation and testing set.

Network	Training set (%)	Validation set (%)	Testing set (%)
ANN	85.60	80.00	75.00
DNN	85.60	85.00	80.00

On the other hand, small data in pharmaceutical research is the key issue to be solved. Although there are many DNN examples about imaging recognition, natural language processing and auto-mobile car, the application of deep-learning in pharmaceutical researches are still very few. Generally speaking, deep-learning methods require a large amount of data for training. This is not a problem in other fields which have the big data source. However, this is a big challenge for the pharmaceutical researches due to the experimental limitation. Thus, the most important problem is how to train a good prediction model on small data with high-dimensions input space. For example, the formulation data of ODTs includes the chemical and physical properties of APIs, multiple excipients with various ratios and four tablets characteristic parameters. In our 145-formulation data, it was found that nearly half of APIs groups' size is less than 3 (small API group). Therefore, the splitting strategy of dataset is critical for model establishment. Firstly, 20-representative testing set was picked up from the whole dataset by pharmaceutical scientists. As for training set and validation set selection, before using automatic selection algorithm, manual selection approach was adopted to ensure the appropriate selection of these two datasets. However, the manual selection needs experts with strong background knowledge, which is time-consuming and non-standardized. When trying the random selection method, the data from small API groups with no representation was easily selected. Thus, the improved maximum dissimilarity algorithm (MD-FIS) is developed to select training set and validation set. MD-FIS is based on the maximum dissimilarity algorithm with the small group filter, representative initial set selection algorithm and new selection cost function. In the MD-FIS process, the data go through a filter to get rid of the data from the small API groups, then the MD-FIS randomly gets the initial datasets, computes each distance from the initial dataset to the corresponding remaining data, and the minimum distance data are chosen as the final initial set. The final initial set and remaining data are the inputs to the dissimilarity algorithm with new selection cost function. The selected data is the validation set, while the remaining data is used as the training set. Because of the small group filter, the validation set from the general groups could represent the feature of whole dataset.

The second important issue is the selection of network algorithm. As deep convolutional networks inspired from visual neuroscience usually achieve a good result for processing images, video, speech and audio [28]. Recurrent neural networks contained history information of the sequence that have brought the breakthrough in sequential data such as text and speech [29]. Our pharmaceutics data only includes properties of API, excipients with its amount and tablet parameters. There is no chronic relationship between each data. Our target is to predict the disintegrating time. Hence, compared with the deep convolutional networks and recurrent neural networks, the full-connected deep feedforward networks should be the best choice for the proposed problem. The challenges about deep feedforward network are computing too many parameters and vanishing the gradient. The results show that the satisfied accuracy could be reached by DNN. The deep-learning method with the proposed data selection algorithms and pharmaceutics evaluation criterion can reach the desired models, which satisfy the accuracy requirements in the pharmaceutics. This deep-learning approach could save a lot of time, manpower and material resource for formulation development of ODTs. This will greatly benefit the formulation design in pharmaceutical research.

Although DNN has reached the expected prediction accuracy on small pharmaceutical datasets, the mechanism of DNN is still a black box, and it is difficult to explain the mapping procedure from the input layer to the output layer. For example, it is unclear how each formulation component contributes to the disintegrating time. Moreover, current model cannot be directly applied to another evaluation parameter of formulations. Current prediction model for ODTs is just the first step in intelligent research for formulation development. Further research in intelligent formulation systems is underway in our laboratory.

4.   Conclusions

The traditional "trial-and-error" method for formulation development has existed hundreds of years, which always costs a large amount of time, financial and human resources. Oral disintegrating tablets are a novel and important formulation form in recent years because of its convenience and good disintegration ability. Current research developed the DNN with MD-FIS select algorithm to establish a good prediction model for the disintegrating time of ODT formulations. On the other hand, this research is also a good example for deep-learning on small data. The proposed predictive approach not only contains formulation information of ODTs, but considering the influence of tablet characteristic parameters, which could evaluate critical parameters of formulation quality and guide formulation research. This deep-learning model could also be applied to other dosage forms and more fields in pharmaceutical research. The implementation of this prediction model could effectively reduce drug product development timeline and material usage, and proactively facilitate the development of a robust drug product.

Conflicts of interest

The authors declare that there are no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Acknowledgement

Appendix.   Supplementary material

Supplementary data to this article can be found online at doi: 10.1016/j.ajps.2018.01.003.

REFERENCES

[1] Bhowmik D, Chiranjib B, Krishnakanth P, Chandira RM. Fast dissolving tablet: an overview. J Chem Pharm Res 2009;1:163–77.

[2] Bandari S, Mittapalli RK, Gannu R. Orodispersible tablets: an overview. Asian J Pharm 2014;2.

[3] Lindgren S, Janzon L. Dysphagia: prevalence of swallowing complaints and clinical finding. Med Clin North Am 1993;77:3–5.

[4] Dutta S, De PK. Formulation of fast disintegrating tablets. Int J Drug Formul Res 2011;201:1.

[5] Fu Y, Yang S, Jeong SH, Kimura S, Park K. Orally fast disintegrating tablets: developments, technologies, taste-masking and clinical studies. Crit Rev Ther Drug Carrier Syst 2004;21.

[6] Prateek S, Ramdayal G, Kumar SU, Ashwani C, Ashwini G, Mansi S. Fast dissolving tablets: a new venture in drug delivery. Am J PharmTech Res 2012;2:252–79.

[7] Kaur T, Gill B, Kumar S, Gupta G. Mouth dissolving tablets: a novel approach to drug delivery. Int J Curr Pharm Res 2011;3:1–7.

[8] Douroumis D. Orally disintegrating dosage forms and taste-masking technologies; 2010. Expert Opin Drug Deliv 2011;8:665–75.

[9] Shukla D, Chakraborty S, Singh S, Mishra B. Mouth dissolving tablets I: an overview of formulation technology. Sci Pharm 2009;77:309–26.

[10] Siddiqui MN, Garg G, Sharma PK. Fast dissolving tablets: preparation, characterization and evaluation: an overview. Int J Pharm Rev Res 2010;4:87–96.

[11] Al-Khattawi A, Mohammed AR. Compressed orally disintegrating tablets: excipients evolution and formulation strategies. Expert Opin Drug Deliv 2013;10:651–63.

[12] Aguilar-Díaz JE, García-Montoya E, Pérez-Lozano P, Suñe-Negre JM, Miñarro M, Ticó JR. The use of the SeDeM Diagram expert system to determine the suitability of diluents–disintegrants for direct compression and their use in formulation of ODT. Eur J Pharm Biopharm 2009;73:414–23.

[13] Pérez P, Suñé-Negre JM, Miñarro M, et al. A new expert systems (SeDeM Diagram) for control batch powder formulation and preformulation drug products. Eur J Pharm Biopharm 2006;64:351–9.

[14] Suñe-Negre J, Roig-Carreras M, Fuster-García R, et al. Nueva metodología de preformulaciÓn galénica para la caracterizaclÓn de sustancias en relaciÓn a su viabilidad oara la comoresiÓn: Diaqrama SeDeM. Cienc Tecnol Pharm 2005;15:125–36.

[15] Suñé-Negre JM, Pérez-Lozano P, Miñarro M, et al. Application of the SeDeM diagram and a new mathematical equation in the design of direct compression tablet formulation. Eur J Pharm Biopharm 2008;69:1029–39.

[16] Aguilar-Díaz JE, García-Montoya E, Suñe-Negre JM, Pérez-Lozano P, Miñarro M, Ticó JR. Predicting orally disintegrating tablets formulations of ibuprophen tablets: an application of the new SeDeM-ODT expert system. Eur J Pharm Biopharm 2012;80:638–48.

[17] Aguilar-Díaz JE, García-Montoya E, Pérez-Lozano P, Suñe-Negre JM, Miñarro M, Ticó JR. SeDeM expert system a new innovator tool to develop pharmaceutical forms. Drug Dev Ind Pharm 2014;40:222–36.

[18] Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Spin glass theory and beyond: an introduction to the replica method and its applications. World Scientific; 1987. p. 411–15.

[19] Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015;61:85–117.

[20] Rost B, Sander C. Combining evolutionary information and neural networks to predict protein secondary structure. Proteins 1994;19:55–72.

[21] Akseli I, Xie J, Schultz L, et al. A practical framework toward prediction of breaking force and disintegration of tablet formulations using machine learning tools. J Pharm Sci 2017;106:234–47.

[22] Dudek AZ, Arodz T, Gálvez J. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. Comb Chem High Throughput Screen 2006;9:213–28.

[23] Murcia-Soler M, Pérez-Giménez F, García-March FJ, et al. Artificial neural networks and linear discriminant analysis: a valuable combination in the selection of new antibacterial compounds. J Chem Inf Comp Sci 2004;44:1031–41.

[24] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.

[25] Xu Y, Dai Z, Chen F, Gao S, Pei J, Lai L. Deep learning for drug-induced liver injury. J Chem Inf Model 2015;55:2085–93.

[26] Baskin II, Winkler D, Tetko IV. A renaissance of neural networks in drug discovery. Expert Opin Drug Discov 2016;11:785–95.

[27] Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V. Deep neural nets as a method for quantitative structure–activity relationships. J Chem Inf Model 2015;55:263–74.

[28] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Proc Mag 2012;29:82–97.

[29] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 1994;5:157–66.