proceedings

MDPI

Proceedings

Computational intelligence model of orally disintegrating tablets – an attempt to explain disintegration process

Jakub Szlęk[1]*, Adam Pacławski[1], Natalia Czub[1], Aleksander Mendyk[1]

[1] Department of Pharmaceutical Technology and Biopharmaceutics, Jagiellonian University Medical College, Medyczna 9, 30-688 Kraków, Poland; j.szlek@uj.edu.pl, adam.paclawski@uj.edu.pl, natalia.czub@doctoral.uj.edu.pl, aleksander.mendyk@uj.edu.pl

* Correspondence: j.szlek@uj.edu.pl; Tel.: +48 12 620 56 04

Abstract: We obtained curated database based on the database presented by Han et al. [1]. Chemical descriptors were introduced as characteristics of active pharmaceutical ingredients (APIs). We used H2O AutoML platform [2, 3] in order to develop a Deep Learning model and SHAP method to explain its predictions [4]. Obtained results were satisfactory with NRMSE of 8.1% and R2 of 0.84. Finally, we identified critical parameters affecting the process of disintegration of directly compressed ODTs.

Keywords: machine learning model; computational intelligence; AutoML; orally disintegrating tablets; ODTs; disintegration time

Citation: Jakub Szlęk*, Adam Pacławski, Natalia Czub, Aleksander Mendyk. Computational intelligence model of orally disintegrating tablets – an attempt to explain disintegration process. *Proceedings* **2021**, *68*, x. https://doi.org/10.3390/xxxxx

Published: date

1. Introduction

Traditional tablets are not an ideal drug dosage form. Many groups of patients, e.g., pediatric or geriatric patients have problems with swallowing or simply are not willing to take tablets. As a consequence, all these factors may reduce patient's compliance. In order to overcome inconvenience of conventional tablet use, orally disintegrating tablets (ODTs) were introduced into the drug market. One of the methods of preparing ODTs is direct compression, which is cost-efficient and simple. It involves comparatively fewer stages than compression preceded by wet or dry granulation. In brief, powders are grinded if necessary and blended, then the mixture is compressed into the tablets. Although the process is quite simple, there are many factors that influence the characteristics of the ODTs, among which one of the crucial factors is the disintegration time.

One of the methods used to solve problems with many factors, where the hypothesis governing the phenomenon is unknown or the whole process is complex, is machine learning (ML). Automated machine learning (AutoML) is currently in focus branch of ML automating the time-consuming, iterative tasks of model development. AutoML enables machine-driven building of large-scale, high-performance, and superb predictability models with minimum human intervention.

Motivation of this study is a limited knowledge of relationships between excipients, APIs and process parameters of direct compression and their influence on disintegration of ODTs. Knowing such behavior would enhance the design and development of a novel drug dosage forms. In this work we applied a concept of AutoML-based heuristic model development for prediction of disintegration time based on the quantitative and qualitative composition of powder mixtures.

2. Materials and Methods

Our database was built based on the database presented by Han et. al [1]. First, we curated the existing database [1] neglecting any unclear or uncertain data records. We put emphasis on the occurrence of the ODTs characteristic and process parameters, such as

tablet hardness, thickness and dimension of tablet press die. Moreover, we performed a literature survey in order to enhance the database. Scopus® database was searched for publications fulfilling following criteria: direct compression method of ODTs should be used in processing, amount of all excipients should be present, tablet characteristics (hardness, thickness, and die dimension) should be present and compendial disintegration test should be applied (Ph. Eur. or USP).

After data scrapping, we calculated APIs two-dimensional (2D) molecular descriptors using mordred-descriptor v.1.2.1a1 Python package [5] and included in the curated database. Excipients types and amounts were encoded in a topological manner. The only output was the time needed for disintegration of tablets.

Computational experiment was performed according to the scheme presented in Figure 1. In brief, preprocessed database was passed to the Python script [2] performing at first stage feature selection, and then final model building according to 10-fold cross validation scheme. All available algorithms in H2O implementation of AutoML were used: Distributed Random Forest (DRF), Extremely Randomized Trees (XRT), Generalized Linear Model (GLM), Extreme Gradient Boosting Machine, (XGBoost), Gradient Boosting Machine (GBM), Deep Learning (fully connected multi-layer artificial neural network), and Stacked Ensemble models.

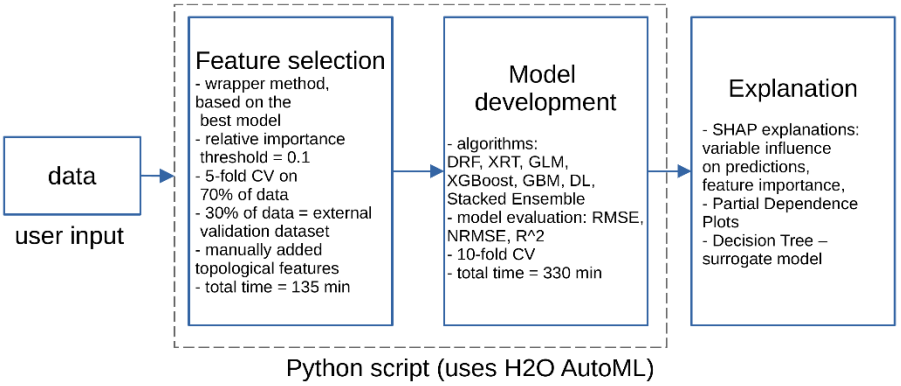


Figure 1. Scheme of computational experiment design.

Model performance was assessed according to the 10-fold cross-validation (10-CV) and expressed by three goodness of fit metrics: root-mean-square error (RMSE), normalized root-mean-square error (NRMSE) and coefficient of determination (R^2). For reference, please see Equation 1, 2 and 3.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (pred_i - obs_i)^2}{n}}, \quad (1)$$

where: obs_i, $pred_i$ = observed and predicted values, i = data record number, n = total number of records.

$$NRMSE = \frac{RMSE}{obs_{max} - obs_{min}} \cdot 100\,\%, \quad (2)$$

where: RMSE = root-mean-square error, obs_{max}, obs_{min} = observed minimal and maximal values.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n} (pred_i - obs)^2}{\sum_{i=1}^{n} (obs_i - obs)^2}, \quad (3)$$

where: R^2 coefficient of determination, SS_{res} = the sum of squares of the residual errors, SS_{tot} = the total sum of the errors, obs_i, $pred_i$ = observed and predicted value, obs - arithmetical mean of observed values.

Predictions of the best model were explained with the use of another Python wrapper [6] implementing among others SHapley Additive exPlanations (SHAP) method by Lundberg et al. [4].

3. Results

Each record of curated database represented one formulation of ODTs. It consisted of 633 chemical descriptors encoding API, 28 inputs encoding amounts of excipients, 9 inputs characterizing drug dosage form. A single independent variable was disintegration time. The database consisted of 243 records (formulations), of which only 52 records (~ 21%) overlapped the Han et al. database [1].

In the feature selection stage inputs number was reduced to 39, among which there were 28 inputs (amount of: 2-hydroxypropyl-beta-cyclodextrin [%], Aerosil [%], Amberlite IRP 64-69 [%], API [%], beta-cyclodextrin [%], calcium silicate [%], camphor [%], colloidal silicon dioxide [%], croscarmellose sodium [%], crospovidone [%], cyclodextrin methacrylate [%], Eudragit EPO [%], hydroxy propyl methyl cellulose [%], lactose [%], low-substituted hydroxy propyl cellulose [%], magnesium stearate [%], mannitol [%], microcrystalline cellulose [%], Poloxamer 188 [%], polyvinyl alcohol [%], polyvinylpyrrolidone [%], pregelatinized starch [%], sodium bicarbonate [%], sodium carboxymethyl starch [%], sodium lauryl sulphate [%], sodium starch glycolate [%], sodium stearyl fumarate [%], and talc [%]) responsible for encoding quantity of excipients and API, 8 molecular descriptors characterizing API (API Geary autocorrelation of lag 7 weighted by ionization potential, API topological charge index of order 7, API Geary autocorrelation - lag 7 / weighted by polarizabilities, API modified information content index, API Moran autocorrelation of lag 4 weighted by polarizability, API negative logarithm of the partition (oil/water) coefficient, and API number of 12-membered rings (includes counts from fused rings), API number of 8-membered fused rings containing heteroatoms (N, O, P, S, or halogens)) and 3 inputs characterizing drug dosage form (diameter of die or tablet [mm], hardness of ODT [N], thickness of ODT [mm]). A list of selected features along with their type and relative importance is presented in Table 1. The full list with relative importance is attached in supplementary material S1.

Table 1. First fifteen selected features and their relative importance.

Feature	Type	Relative importance
CC_Na_perc	Amount of croscarmellose sodium [%]	1.0000
Crospovidone_perc	Amount of crospovidone [%]	0.8013
SSG_perc	Amount of sodium starch glycolate [%]	0.7341
Hardness_N	Hardness of ODT [N]	0.6564
Eudragit_EPO_perc	Amount of Eudragit EPO [%]	0.5620
Mg_stearate_perc	Amount of magnesium stearate [%]	0.5008
Aerosil_perc	Amount of Aerosil [%]	0.3991
GATS7i	API Geary autocorrelation of lag 7 weighted by ionization potential	0.3441

MCC_perc	Amount of microcrystalline cellulose [%]	0.3394
Colloidal_silicon_dioxide_perc	Amount of colloidal silicon dioxide [%]	0.2336
Mannitol_perc	Amount of mannitol [%]	0.2335
Pregelatinized_starch_perc	Amount of pregelatinized starch [%]	0.2009
PVA_perc	Amount of polyvinyl alcohol [%]	0.1618
Thickness_mm	Thickness of ODT [mm]	0.1482
CD_methacrylate_perc	Amount of cyclodextrin methacrylate [%]	0.1253
(...)	(...)	(...)
Disintegration_time_sec	Disintegration time [s]	**output**

The best results were obtained by a Deep Learning (DL) model, which had RMSE = 10.9, NRMSE = 8.1% and R^2 = 0.84. The model had 2 hidden layers with 100 neurons in each layer and rectifier with dropout as an activation function. A plot of predicted versus observed disintegration values is presented in Figure 2.

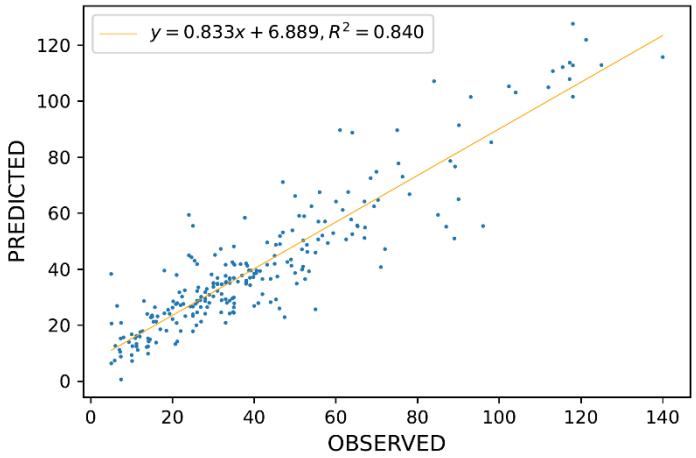
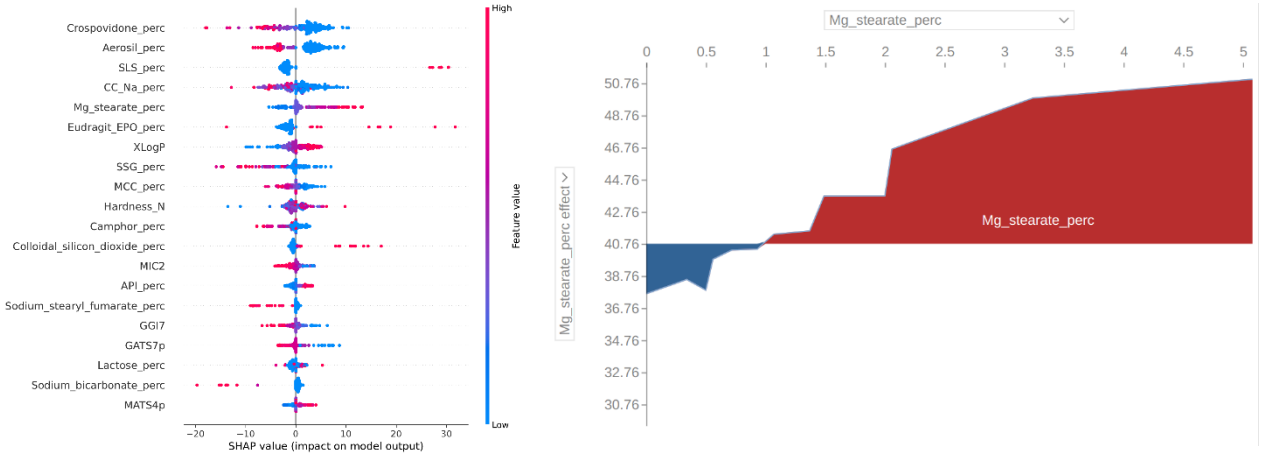


Figure 2. Predicted vs. Observed values for disintegration time for Deep Learning model.

Following model development, a procedure of SHAP method was applied. Then selected plots were analyzed, and conclusions were drawn (Figure 3).



(**a**) (**b**)

Figure 3. Results of model's explanation: (**a**) Summary plot of impact on model output and feature value; (**b**) Effect of magnesium stearate amount [%] on average model's prediction.

4. Discussion

Based on the obtained prediction metrics (RMSE, NRMSE, R^2), it can be concluded that the model is satisfactory in terms of generalization. The 10-fold cross validation technique was used as a golden standard. The mean error of the model is 10.9 (NRMSE = 8.1%), therefore it is possible to optimize a formulation with its use. Moreover, in Figure 3a, critical parameters and their impact on disintegration time were identified. It seems that, high amount of sodium lauryl sulphate, magnesium stearate, Eudragit EPO, colloidal silicon dioxide could increase disintegration time of ODTs. On the other hand, high amount of: crospovidone, Aerosil, croscarmellose, sodium starch glycolate, or sodium stearyl fumarate could lead to decreased disintegration time. Looking more closely at the variable effects a percolation threshold could be pointed. For example, at Figure 3b at magnesium stearate value of about 1% a revers in effects could be observed. This observation is consistent with finding of previous studies [7]. It is believed that magnesium stearate in higher amounts than 1%, beside usual action as lubricant, could form hydrophobic film around API particles and could prevent water from penetrating into the core of the tablet. Using the similar reasoning, the XLogP limit was determined for the API, the value of which will increase the disintegration time of ODTs (Figure 4). The general conclusion is that more hydrophobic API with XLogP higher than 3.5 would negatively affect the disintegration time increasing it.

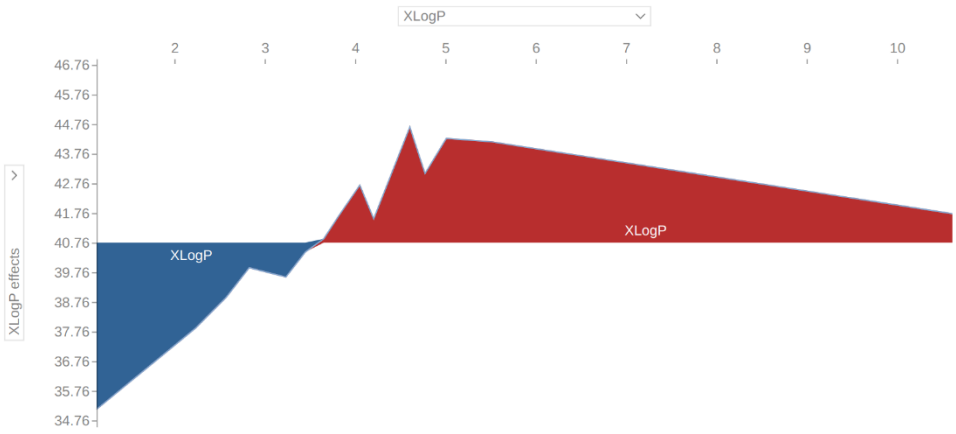


Figure 4. Effect of XLogP (calculated partition coefficient of API) on average model's prediction.

Author Contributions: Conceptualization, J.S. and A.M..; methodology, J.S., A.P., N.Cz., A.M.; software, J.S.; validation, J.S., A.P., N.Cz., A.M.; formal analysis, J.S.; investigation, J.S.; resources, A.M.; data curation, J.S.; writing—original draft preparation, J.S., A.P., N.Cz., A.M.; writing—review and editing, J.S., A.P., N.Cz., A.M.; visualization, J.S.; supervision, J.S.; project administration, J.S.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Uniwersytet Jagielloński Collegium Medicum, grant number N42/DBS/000205.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available upon written request.

Acknowledgments: Not applicable.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. The authors declare no conflict of interest.

References

1. Han R.; Yang Y.; Li X.; Ouyang D. Predicting oral disintegrating tablet formulations by neural network techniques. Asian J. Pharm. Sci., **2018**, 134, 336-342.
2. Szlęk J. h2o_AutoML_Python, Python script for AutoML in h2o. Available online: https://github.com/jszlek/h2o_AutoML_Python (accessed on 10 August 2021).
3. LeDell, E.; Poirier, S. H2O AutoML: Scalable Automatic Machine Learning. 7th ICML Workshop on Automated Machine Learning. **2020**.
4. Lundberg, S. M.; Erion, G.G; Lee S.I. Consistent Individualized Feature Attribution for Tree Ensembles. arXiv preprint arXiv:1802.03888. **2018**.
5. Moriwaki H.; Tian Y.-S.; Kawashita N.; Takagi T.; Mordred: a molecular descriptor calculator. Journal of Cheminformatics **2018,** 10:4, 1-14. doi: 10.1186/s13321-018-0258-y.
6. Szlęk J. **2021**. Model Interpretation. Available online: https://github.com/jszlek/MODEL_INTERPRETATION (accessed on 10 August 2021).
7. Jennifer Wang, Hong Wen, Divyakant Desai, Lubrication in tablet formulations, European Journal of Pharmaceutics and Biopharmaceutics, Volume 75, Issue 1, **2010**, Pages 1-15, ISSN 0939-6411, https://doi.org/10.1016/j.ejpb.2010.01.007.