Macedonian pharmaceutical bulletin, 69 (Suppl 1) 113- 114 (2023)
Online ISSN 1857 - 8969
DOI: 10.33320/maced.pharm.bull.2023.69.03.055
Short communication

Application of the Gradient boosted tree approach for thin film classification based on disintegration time

Erna Turković[1], Ivana Vasiljević[1], Jelena Parojčić[1]

[1]*Department of Pharmaceutical Technology and Cosmetology, University of Belgrade-Faculty of Pharmacy, Vojovde Stepe 450, 1100 Belgrade, Serbia*

Introduction

Thin films are polymeric strips that disintegrate in the oral cavity and consist of a film-forming agent and an active pharmaceutical ingredient (API). Generally, thin films disintegrate within seconds, but their composition can be modified to allow slower disintegration and release of the loaded API, depending on the properties of the film. Research into various aspects of oral thin films is progressing rapidly, but thin films are also being discussed in the context of a broader range of other dosage forms, such as carrier for multiparticulates or nano-based dosage forms and for the fixed-dose combinations (Turković et al., 2022). Large amounts of data are being generated over the years, so integrating machine learning algorithms can be beneficial to gain more in-depth knowledge about the thin film properties and interactions between film constituents. Gradient boosted tree is one of machine learning tools that perform regression or classification by combining the outputs from individual decision trees. This work is aimed to explore the possibility of integrating a machine learning approach in evaluation of experimental data obtained by films characterization. Potential application of Gradient boosted trees for thin films characterization based on their disintegration properties as film critical quality attribute was investigated.

Materials and methods

Materials

Eight hydrophilic polymers were investigated as film-forming agents: (1) hydroxypropyl cellulose (Klucel® GF, Ashland™, USA), (2) hypromellose (Pharmacoat 606, Shin-Etsu Chemical Co., Japan); (3) carboxymethylcellulose sodium salt (Fluka Chemie AG, Switzerland), (4) polyethylene glycol–polyvinyl alcohol graft copolymer (Kollicoat® IR, BASF, Germany), (5) maltodextrin (Glucidex IT12, Roquette, France), (6) sodium alginate (Fisher Scientific, USA), (7, 8) poly(ethylene oxide) polymers (POLYOX™ WSR N10, PEO N80, POLYOX™ WSR N80, DuPont, U.S.).

Glycerol (Gly), 85% (w/w) (Ph.Eur) was used as plasticizer, and magnesium aluminometasilicate (Neusilin UF, Fuji Chemical Industries Co, Japan), croscarmellose sodium (Primellose®, DFE Pharma, Germany), crospovidone (Polyplasdone™ XL-10, Ashland™, USA), sodium starch glycolate (SSG, Primojel®, DFE Pharma, Germany), calcium silicate (CaS, RxCIPIENTS® FM1000, Huber Engineered Materials, Havre de Grace, MD, USA) were used as disintegrants. Active pharmaceutical ingredients which were used are ibuprofen, paracetamol, caffeine, enalapril, verapamil, atenolol, carvedilol (Ph. Eur).

Sample characterization

Samples were evaluated with respect to weight and thickness uniformity, disintegration time and mechanical properties, including oscillatory rheology for complex modulus assessment (Drašković et al., 2020).

Machine learning model development

RapidMiner 9.10.011 software (RapidMiner Studio, Massachusetts, USA) was used for model development. A model was built starting from creating validation and training sets. Stratified sampling was employed where random subsets were prepared and the class distribution in the subsets was the same as in the whole dataset. Some

*erna.turkovic@pharmacy.bg.ac.rs

missing values were detected for one attribute and were replaced with the average numerical values. After those steps, actual training was performed, and cross validation was employed for the generated model.

Results and discussion

The investigated dataset included main formulation factors, i.e. the selected API, polymer, disintegrant, plasticizer and their concentration, whether API was dissolved or dispersed in the medium and molecular weight of the polymers. The obtained experimental results which included film weight, thickness, complex modulus, tensile strength, elongation at break, Young's modulus were also included as input, while disintegration time was the explored outcome. The target classes were disintegration times (DT) shorter than 30 or 60 seconds, or longer than 60 seconds.

The outlier operator was able to identify one outlier among the hundred samples, which showed exceptionally different mechanical properties compared to the samples with the same polymer ratio and type, so ninety-nine samples were used for model development.

The Gradient boosted tree was developed by training one tree at a time, each tree correcting the errors of the previous one so that the total number of generated trees was 90, while the number of internal trees was 270.

Table 1. Confusion matrix

	true DT<30	true DT<60	true DT>60	class precision (%)
pred. DT<30	10	1	0	90.91%
pred. DT<60	3	10	0	76.92%
pred. DT>60	1	0	4	80.00%
class recall (%)	71.43%	90.91%	100.00%	

In Table 1, the 3 x 3 confusion matrix is shown to describe the performance of a model. True/false positive and true/false negative values were detected. The class precision value is highest for the pred. DT<30 class, indicating that it had the highest proportion of matching data among the retrieved data. The class recall, meaning the proportion of relevant data that was retrieved, is highest for the DT>60 class. The overall model had an accuracy of 82.67 ± 1.49%, indicating a high number of correct classifications achieved with model.

Figure 1 shows the obtained attributes weight. The polymer type and the selected API had the highest weights, indicating that these two parameters were the most important contributors to the model. Further analysis is needed to evaluate whether some of the attributes can be excluded from the database so that the time required for data collection can be reduced. The polymer type was the input with the highest value of variable relative importance (67.51) which indicates that the polymer type variable was most frequently selected to split on during the tree building and the squared error improved as a result.

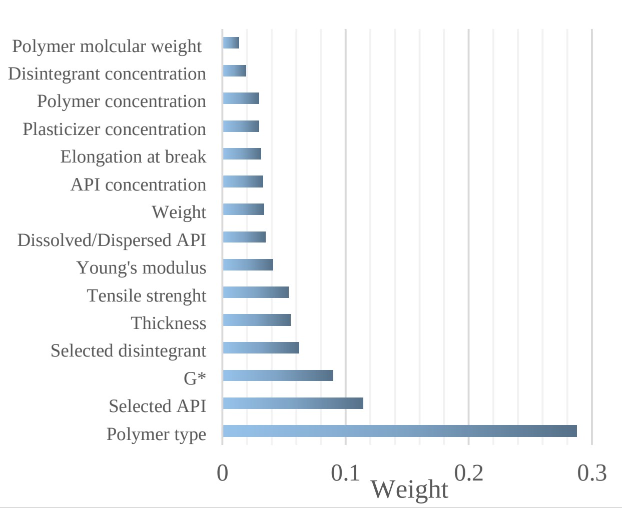

Fig. 1. Attributes weight

Conclusion

The obtained results indicate that Gradient boosted tree algorithm can be employed to accurately classify thin films based on their disintegration time. Polymer type was identified as critical variable affecting thin film disintegration time, and predictive model development.

This work is a screening study that demonstrates that machine learning can be a valuable tool for pharmaceutical application, as it can potentially facilitate development of dosage forms with targeted quality attributes.

Acknowledgment: This work was financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (project 451-03-68/2022-14/200161).

References

Drašković, M., Turković, E., Vasiljević, I., Trifković, K., Cvijić, S., Vasiljević, D., Parojčić, J., 2020. Comprehensive evaluation of formulation factors affecting critical quality attributes of casted orally disintegrating films. J. Drug Deliv. Sci. Technol. 56, 101614. https://doi.org/10.1016/j.jddst.2020.101614

Turković, E., Vasiljević, I., Drašković, M., Parojčić, J., 2022. Orodispersible films — Pharmaceutical development for improved performance: A review. J. Drug Deliv. Sci. Technol. 75. https://doi.org/10.1016/j.jddst.2022.103708