European Journal of Pharmaceutical Sciences 124 (2018) 295–303

Contents lists available at ScienceDirect

European Journal of Pharmaceutical Sciences

journal homepage: www.elsevier.com/locate/ejps

Optimization of pellets manufacturing process using rough set theory

Łukasz Pałkowski[a,*], Maciej Karolak[a], Bartłomiej Kubiak[a,b], Jerzy Błaszczyński[c], Roman Słowiński[c,d], Markus Thommes[e], Peter Kleinebudde[f], Jerzy Krysiński[a]

[a] Nicolaus Copernicus University, Collegium Medicum, Department of Pharmaceutical Technology, Jurasza Str. 2, 85-094 Bydgoszcz, Poland
[b] Adamed Group, Pieńków 149, 05-152 Czosnów near Warsaw, Poland
[c] Poznań University of Technology, Institute of Computing Science, Piotrowo Str. 2, 60-965 Poznań, Poland
[d] Systems Research Institute, Polish Academy of Sciences, Newelska Str. 6, 01-447 Warszawa. Poland
[e] Laboratory of Solids Process Engineering, Technical University of Dortmund, Emil-Figge-Str. 68, 44227 Dortmund, Germany
[f] Institute of Pharmaceutics and Biopharmaceutics, Heinrich-Heine-University, 40225 Düsseldorf, Germany

ARTICLE INFO

Keywords:
Aspect ratio
Pellets
Knowledge discovery
Rough set theory

ABSTRACT

Pharmaceutical pellets are spherical agglomerates manufactured in extrusion/spheronization process. The composition of the pellets, the amount of active pharmaceutical ingredient (API) and the type of used excipients have an influence on the shape and quality of dosage form. A proper quality of the pellets can also be achieved by identifying the most important technological process parameters. In this paper, a knowledge discovery method, called dominance-based rough set approach (DRSA) has been applied to evaluate critical process parameters in pellets manufacturing. For this purpose, a set of condition attributes (amount of API; type and amount of excipient used; process parameters such as screw and rotation speed, time and temperature of spheronization) and a decision attribute (quality of the pellets defined by the aspect ratio) were used to set up an information system. The DRSA analysis allowed to induce decision rules containing information about process parameters which have a significant impact on the quality of manufactured pellets. Those rules can be used to optimize the process of pellets manufacturing.

1. Introduction

Methods of knowledge discovery and data mining are increasingly exploited in pharmaceutical technology for process and product optimisation (Gardiner and Gillet, 2015). These methods are applied for both, development of a new product, and quality control or risk management of existing products. The process analytical technology (PAT) initiative and quality by design (QbD) approach to product design and production (ICH Q8–Q10) have been recently incorporated by the European Medicines Agency and the US Food and Drug Administration (Yu, 2008). These methods ensure higher quality products and faster development. Basically, the framework aims at a better process understanding and process design to ensure a required quality. Different methods have been applied to perform these analysis such as statistical tools (DoE—Design of Experiment), data mining and computational intelligence methods, like Artificial Neural Networks (ANNs), decision trees, genetic algorithms, and fuzzy logic (Lourenço et al., 2011). Simple DoE methods have some disadvantages. Underlying the use of two-level factors is the belief that the resultant changes in the dependent variable are basically linear in nature. However, this is often not true, because many variables are related to quality characteristics in a non-linear manner. Fractional design problem is the implicit assumption that higher-order interactions do not matter. When some attributes are set to a particular level, one attribute may be negatively related to product quality. Again, in fractional factorial designs, in particular higher-order interactions (greater than two-way), will escape detection. DoE methods are appropriate for construction of a prognostic model, however not for cause-effect relationship analysis.

In this paper, we present an application of a novel knowledge discovery technique, which overcomes the above mentioned disadvantages, and it allows to handle a mixture of discrete and continuous data. The presented method is based on rough set theory (RST).

Dominance-based rough set approach (DRSA) is an implementation of RST adapted to ordinal data. It was chosen as the most suitable method to discover synthetic rules that exhibit monotonic relationships between composition and process parameters of pellets on the one hand, and their final quality on the other hand. It is well suited for this application because it handles qualitative and quantitative attributes, without the need of discretization of quantitative attributes or transformation of qualitative attributes into quantitative ones. DRSA is also

* Corresponding author.
E-mail address: lukaszpalkowski@cm.umk.pl (Ł. Pałkowski).

https://doi.org/10.1016/j.ejps.2018.08.027
Received 23 February 2018; Received in revised form 3 July 2018; Accepted 22 August 2018
Available online 26 August 2018

able to deal with possible inconsistencies in the information table prior to induction of rules. Moreover, it handles global or local monotonic relationships between values of condition attributes and the quality classes (Greco et al., 2001). These characteristics of DRSA perfectly fit the type of data to be analyzed. Using this method, we obtain decision rules with ranges of values of condition attributes in particular classes of pellet's quality.

Pellets are spherical agglomerates in a range between 0.5 and 2.0 mm. Their optimal shape and reproducible surface make them ideal for coating and multiparticulate applications (Swarbrick, 2007). Wet extrusion/spheronization (E/S) is an established and widespread technique to produce pharmaceutical pellets. Pellet preparation using this method is based on 3 different processes: extrusion, spheronization, and fluid bed drying. Parameters affecting pellet properties are the type of excipients: microcrystalline cellulose (MCC), carrageenan, starch, lactose; the water content of extrudate, the physicochemical properties of active pharmaceutical ingredients (API) and technological parameters (screw speed, the number of die holes, spheronizer speed, the time of spheronization, spheronizer temperature) (Fekete et al., 1998). The shape of pellets can be described by a parameter called aspect ratio (AR); the ratio of the maximum length of pellets divided by the orthogonal width. To evaluate the relationship between the parameters describing the manufacturing process and properties of the product chemometric, data analysis methods are useful. RST is a mathematical approach to knowledge discovery from imperfect data, i.e., affected by partial inconsistency. DRSA extends the original RST on inconsistency not only with respect to indiscernibility, but also with respect to dominance, which permits analysis of ordinal data (Błaszczyński et al., 2012). With artificial neural networks and statistical methods, DRSA constitute data analysis tools useful for knowledge discovery from data. RST has been successfully applied in medical area, e.g., for assessing preterm birth risk (Tsumoto, 2000), diagnosis of acute appendicitis (Wilk et al., 2005), breast cancer treatment (Załuski et al., 2004; Chen et al., 2011), prediction of heart attacks (Srinivas et al., 2014), diagnosis of diabetes (Margret Anouncia et al., 2013) and tuberculosis (Uçar et al., 2013). There are some pharmaceutical applications, i.e. structure-activity relationships analysis of antifungal activity imidazolium compounds (Pałkowski et al., 2015), antibacterial activity quaternary ammonium chlorides (Pałkowski et al., 2014; Pałkowski et al., 2012), and fluoroquinolones (Liu et al., 2007).

The aim of this study was to discover critical attributes (process parameters) of pellet formulation which significantly affect the shape of the pellets, and to determine the impact of carrageenan on pellet formulation process as an alternative to using of MCC, allowing to avoid problems with disintegration of the pellets. Moreover, this study was to discover relationships between pellet properties (formulation, technological conditions) and product quality. This is the first application of DRSA in pharmaceutical technology.

2. Material and methods

2.1. Data set

Data for the analysis describing the parameters of production and the quality of the pellets come from the Kleinebudde team published previously (Thommes and Kleinebudde, 2007; Bornhöft et al., 2005; Thommes et al., 2007; Thommes and Kleinebudde, 2006a; Thommes and Kleinebudde, 2006b; Thommes and Kleinebudde, 2008). Analyses were related to the effect of various excipients and technological parameters on the shape of pellets. The formulations of the pellets were characterized by conditional attributes describing quantitative composition: the amount of API (acetaminophen, theophylline, mesalazine, hydrochlorothiazide, phenacetin, chloramphenicol, dimenhydramin, lidocaine); amount of excipient (lactose, mannitol, starch, dicalcium phosphate, MCC and κ-carrageenan (Satiagel® CT 27 — Degussa, Germany; Gelcarin® GP 812 NF (Gel812); Gelcarin® GP 911 NF (Gel911) — FMC, USA; Genugel® X-930-03 — CPCelco, Denmark). Another conditional attributes were API properties (logP, solubility); manufacturing process parameters: extrusion (screw speed, the number of die holes), spheronization (rotation speed, time, temperature), drying (temperature); loss on drying. Decision attribute classifying objects was aspect ratio parameter (AR). Table 1 presents condition attributes used in the information system.

Table 1
Condition attributes and their domains.

Condition attribute	Domain
API [%]	10–95
LogP of API	−0.4–1.6
API solubility [g/l]	0.7–700
Lactose [%]	0, 20, 40, 60
Mannitol [%]	0, 20, 40, 60
Starch [%]	0, 20, 40, 60
$CaHPO_4$ [%]	0, 20
Satiagel [%]	0, 20
Gel812 [%]	0, 20
Gel911 [%]	0, 5, 10, 20
Genugel [%]	0, 20
MCC [%]	0, 20
Screw speed [rpm]	50, 100, 125, 200
Number of die holes	3, 13, 23
Rotation speed [rpm]	500, 750, 1000
Spheronization temperature [°C]	15, 25, 30, 45
Time of spheronization [s]	15–300
Temperature of drying [°C]	60, 105
Loss on drying [%]	31.77–125

The parameters describing spheronization conditions, namely spheronization time, speed, and temperature, screw speed, and number of die holes were optimized in previous experiments by Kleinebudde et al. (Thommes and Kleinebudde, 2007; Bornhöft et al., 2005; Thommes et al., 2007; Thommes and Kleinebudde, 2006a; Thommes and Kleinebudde, 2006b; Thommes and Kleinebudde, 2008). The values of those parameters presented in the analyzed information system appeared to be robust regarding the aspect ratio of the pellets.

In the course of analysis, AR values were discretized as follows:

– optimal (spherical) pellets, called class 1: AR $\leq$ 1.1,
– improper (irregular) pellets, called class 2: AR $>$ 1.1.

The above discretization comes from an observation (Thommes and Kleinebudde, 2006a) that pellets with an optimal shape possess AR value lower or equal to 1.1, while pellets with AR value $>$ 1.1 are improper.

2.2. Information system

The data set analyzed using DRSA is organized in the tabular form of an information system, where a set of objects (products = pellets) is described by a finite set of condition attributes and one decision attribute (quality class of the pellets). Rows of such a table correspond to objects and columns to attributes, and at the intersection of rows and columns there are values called descriptors.

Table 2 presents a part of the information system describing a set of different formulations of pellets. The whole information system was built based on 227 pellets formulations and can be found in the Supplementary material.

2.3. Decision rules

Decision rules represent important cause-effect relationships between values of condition and decision attribute. Rules consist of condition and decision parts, called premise and conclusion, respectively. The rules have the following syntax: “if the conjunction of elementary

Table 2
A part of the information system.

No.	API [%]	LogP of API	Solubility [g/l]	Lactose [%]	Mannitol [%]	Starch [%]	$CaHPO_4$ [%]	Satiagel [%]	Gel812 [%]	Gel911 [%]	Genugel [%]
1	80	0.5	14.3	0	0	0	0	0	0	20	0
2	80	0.5	14.3	0	0	0	0	0	0	20	0
3	80	0.5	14.3	0	0	0	0	0	0	0	0
4	80	0.5	14.3	0	0	0	0	0	0	0	0
5	80	0.5	14.3	0	0	0	0	0	0	0	0
6	60	0.5	14.3	20	0	0	0	0	0	20	0
7	60	0.5	14.3	20	0	0	0	0	0	20	0
8	60	0.5	14.3	20	0	0	0	0	0	20	0
9	40	0.5	14.3	40	0	0	0	0	0	20	0
10	40	0.5	14.3	40	0	0	0	0	0	20	0
11	40	0.5	14.3	40	0	0	0	0	0	20	0
12	20	0.5	14.3	60	0	0	0	0	0	20	0
13	20	0.5	14.3	60	0	0	0	0	0	20	0
14	20	0.5	14.3	60	0	0	0	0	0	20	0
15	40	0.5	14.3	40	0	0	0	0	0	0	0
16	40	0.5	14.3	40	0	0	0	0	0	0	0
17	40	0.5	14.3	40	0	0	0	0	0	0	0
18	60	0.5	14.3	0	20	0	0	0	0	20	0
19	60	0.5	14.3	0	20	0	0	0	0	20	0
20	60	0.5	14.3	0	20	0	0	0	0	20	0
21	40	0.5	14.3	0	40	0	0	0	0	20	0
22	40	0.5	14.3	0	40	0	0	0	0	20	0
23	20	0.5	14.3	0	60	0	0	0	0	20	0
24	20	0.5	14.3	0	60	0	0	0	0	20	0
25	20	0.5	14.3	0	60	0	0	0	0	20	0

No.	MCC [%]	Screw speeed [rpm]	Number of die holes	Spheroniza-tion speed [rpm]	Spheroniza-tion tempera-ture[°C]	Time of spheroniza-tion [s]	Temperatu-re of drying [°C]	Loss on drying [%]	AR
1	0	100	23	750	25	300	105	58.70	1.1545
2	0	100	23	750	25	300	105	64.40	1.1221
3	20	100	23	750	25	300	105	56.20	1.0697
4	20	100	23	750	25	300	105	53.00	1.0692
5	20	100	23	750	25	300	105	58.60	1.0681
6	0	100	23	750	25	300	105	85.20	1.0721
7	0	100	23	750	25	300	105	77.50	1.0949
8	0	100	23	750	25	300	105	74.30	1.1454
9	0	100	23	750	25	300	105	74.00	1.0839
10	0	100	23	750	25	300	105	67.00	1.1301
11	0	100	23	750	25	300	105	79.90	1.0878
12	0	100	23	750	25	300	105	75.23	1.1349
13	0	100	23	750	25	300	105	70.09	1.1717
14	0	100	23	750	25	300	105	81.20	1.1177
15	20	100	23	750	25	300	105	48.40	1.0742
16	20	100	23	750	25	300	105	50.50	1.0693
17	20	100	23	750	25	300	105	54.10	1.0784

Table 2 (*continued*)

No.	MCC [%]	Screw speed [rpm]	Number of die holes	Spheronization speed [rpm]	Spheronization temperature [°C]	Time of spheronization [s]	Temperature of drying [°C]	Loss on drying [%]	AR
18	0	100	23	750	25	300	105	79.00	1.0751
19	0	100	23	750	25	300	105	72.00	1.0806
20	0	100	23	750	25	300	105	66.00	1.0859
21	0	100	23	750	25	300	105	71.50	1.0702
22	0	100	23	750	25	300	105	65.60	1.0835
23	0	100	23	750	25	300	105	72.20	1.0669
24	0	100	23	750	25	300	105	65.20	1.0800
25	0	100	23	750	25	300	105	58.90	1.1166

conditions on selected attributes is fulfilled, then the object belongs to some indicated quality class". An elementary condition has one of the forms: "attribute a_i has value at least r_i", "attribute a_i has value at most s_i", "attribute a_i has value equal to t_i", "attribute a_i has value in the interval $[r_i, s_i]$", "attribute a_i takes a value from the finite set $\{r_i, s_i, t_i, \ldots\}$", where r_i, s_i, t_i, … are some values belonging to the domain of attribute a_i, discovered in the induction process. Rules involve only significant condition attributes that have the greatest impact on the decision. Therefore, the set of decision rules is presented in the form of a tabular information system from which unnecessary and redundant information relating to cause-effect relationships in the analyzed data has been removed.

Rules are characterized by their strength defined as a ratio of the number of pellets matching the condition part of the rule to the total number of pellets formulation in the sample. Sets of decision rules, which are essential for the analysis presented in this work, were induced from pellets data, which were collected in an information system. A part of the system can be seen in Table 2.

Decision rules are also characterized by Bayesian confirmation measures, which indicate usefulness of knowledge represented by a premise for a correct classification of pellets to a given class (Greco et al., 2016). It therefore quantifies the degree to which the premise supports the conclusion of the rule.

2.4. Knowledge discovery technique

DRSA analysis was performed using jRS (java Rough Set) library and jMAF (java Multi-criteria and Multi-attribute Analysis Framework) software. Decision rules were generated by VC-DomLEM algorithm (Greco et al., 2001). Considering the order of condition attributes value sets positive or negative monotonic relationship with decision attributes can be distinguished. Positive relationship imply situation where the greater condition attribute value, the better class of decision attribute is achieved (in this case better quality of pellets). Analogously, in negative relationship it is assumed that the smaller value of condition attribute, the more likely it is to obtain improper (irregular) pellets. Some of the attributes have been transformed by duplication of attribute value, i.e., we are considering such attributes in two copies. The first one is assumed to have positive (gain) and the other one to have negative monotonic relationship (cost). The applied transformation of data is non-invasive and does not cause disruption in discovering specific monotonic relationships between condition and decision attributes (Błaszczyński et al., 2009). The rule induction algorithm constructs decision rules involving elementary conditions on one or both copies of particular attributes. For example, in a rule indicating the assignment of a pellet to class 1 there may appear the following elementary conditions concerning attribute a_i:

- $\uparrow a_i(\text{pellet}) \geq \text{val}_{i1}$,
- $\downarrow a_i(\text{pellet}) \leq \text{val}_{i2}$,
- $\uparrow a_i(\text{pellet}) \geq \text{val}_{i1}$ and $\downarrow a_i(\text{pellet}) \leq \text{val}_{i2}$, which boils down to $a_i(\text{pellet}) \in [\text{val}_{i1}, \text{val}_{i2},]$ if $\text{val}_{i1} \leq \text{val}_{i2}$,

where $\uparrow a_i$ and $\downarrow a_i$ are positive (gain) and negative (cost) copies of attribute a_i, respectively. Please note that the transformation of attributes permits discovering global and local monotonic relationships between properties of pellets and their class assignment. The monotonic relationship is global when it can be expressed by a single elementary condition concerning positive or negative attribute. Local monotonicity relationship requires conjunction of two elementary conditions of different type.

2.5. Bayesian confirmation measure s

The effect of condition attributes on predictive accuracy was verified by determining the value of the confirmation measure for each attribute. From among many available Bayesian confirmation measures, we selected measure called *s*, defined as a difference of conditional probabilities Pr(conclusion | premise) − Pr(conclusion | ¬premise). The confirmation measure *s* takes the values between −1 to 1. The higher a positive value of *s*, the greater impact has the premise on the conclusion. Analogically, the negative value of *s* speaks about disconfirmation of the conclusion by the premise. A zero value of *s* means no impact. This relatively simple measure has good mathematical properties, as demonstrated in (Greco et al., 2016).

The confirmation measure *s* is also useful to rank condition attributes according to their impact on the correct classification of pellets (Błaszczyński et al., 2011). In this case, the Bayesian confirmation measure quantifies the degree to which the presence of attribute a_i in premise, denoted by a_i⊳premise, provides evidence for or against correct classification by a rule. Measure *s*(correct, a_i⊳premise) is, in this case, defined as follows, *s*(correct, a_i⊳premise) = Pr(correct|a_i⊳premise) − ⊳Pr(correct|a_i¬⊳premise). In result, attributes present in the premise of a rule which classifies correctly, and attributes absent in premise of a rule which classifies incorrectly, are getting more relevant.

2.6. Stratified cross-validation

Stratified 5-fold cross-validation procedure was used to assess the predictive accuracy of rules. Variable consistency bagging (VC-bagging) (Błaszczyński et al., 2010a; Błaszczyński et al., 2010b) was applied to increase accuracy of results produced by VC-DomLEM. In this procedure, the analyzed data set was divided randomly into training set and a test set in a ratio of 4 to 1. Decision rules constructed on the training set were validated on the test set. The procedure was repeated 1000 times to obtain acceptable reproducibility of results. The final decision rules are the most relevant rules of all the repetitions.

3. Results

Decision rules contain the most important information characterizing pellets in class 1 (optimal shape) and class 2 (inappropriate shape), that is, the characteristics of pellets taking the spherical form (AR ≤ 1.1), or non-spherical form (AR > 1.1). The rules presented in Tables 3 and 4 are ranked according to rule support value. Condition attributes that were not included in decision rules during the analysis were removed from the tables.

Strong decision rules, supported by a large number of objects, with high confirmation measure *s* obtained for class 1 (Table 3) allow to indicate condition attributes and their ranges for spherical pellets. Those features include:

- Rule 1: If API solubility ≤ 0.8 g/l and loss on drying ≥82.7%, then class 1;
- Rule 2: If logP API ≤ 0.5 and amount of starch ≤ 20% and number of die holes ≥ 23 and loss on drying ≥ 70.9%, then class 1;
- Rule 4: If Amount of API ≥ 40.0% and logP API ≤ 0.5 and amount of κ-carrageenan (Genugel 911) ≥ 20.0% and number of die holes ≥ 23 and loss on drying ≥ 74%, then class 1;
- Rule 6: If API solubility ≤ 0.7 g/l and loss on drying 79–109.84%, then class 1;
- Rule 10: If amount of API ≤ 40% and logP API ≤ 0.5 and amount of lactose ≤ 40% and lack of starch and number of die holes ≥ 23, then class 1.

The selected decision rules can be used as a guide to optimize pellet composition and extrusion and spheronization process parameters resulting in optimum shape pellets.

Strong decision rules were also obtained for class 2 of the pellets (Table 4). They provide information about the attributes and their values that adversely affect the shape of the resulting pellets, i.e., the composition and process parameters that are not worth using. Those features include:

- Rule 1: If the screw number of die holes ≤ 13 and rotation speed ≤ 750 rpm and loss on drying ≤ 82.1%, then class 2;
- Rule 2: If API solubility ≤ 0.8 g/l and screw speed ≤ 125 rpm and number of die holes ≤ 13 and rotation speed ≤ 750 rpm and loss on drying ≤ 82.1%, then class 2;
- Rule 3: If solubility ≤ 0.8 g/l and number of die holes ≤ 13 and rotation speed ≤ 750 rpm and spheronization temperature ≤ 30 °C and loss on drying ≤ 82.1%, then class 2
- Rule 4: If logP API ≥ 1.2 and lack of MCC and rotation speed ≤ 750 rpm and loss on drying ≤ 82.1%, then class 2;
- Rule 6: If lack of lactose and mannitol and MCC and rotation speed ≤ 750 rpm and loss on drying ≤ 73.5%, then class 2;
- Rule 11: If logP API ≥ 1.2 and lack of lactose and lack of $CaHPO_4$ and amount of κ-carrageenan (Genugel 911) ≥ 20% and rotation speed ≤ 750 rpm and spheronization temperature ≤ 25 °C, then class 2;

Table 3
Decision rules for class 1 pellets.

No.	API [%]	LogP	Solubility [g/l]	Lactose [%]	Starch [%]	Gel911 [%]	Gel812 [%]	Number of die holes	Loss on drying [%]	Rule support	Rule strength	Confirmation measure *s*
1			≤0.8						≥82.7	50	0.2202	0.52
2		≤0.5			≤20			≥23	≥70.9	46	0.2026	0.51
3			≤0.8						≥85.1	46	0.2026	0.51
4	≥40	≤0.5				≥20		≥23	≥74.0	45	0.1982	0.52
5			≤0.7						82.51–116.05	44	0.1938	0.51
6			≤0.7						79.0–109.84	43	0.1894	0.58
7									85.1–97.18	43	0.1894	0.51
8			≤0.7						82.51–106.35	40	0.1762	0.51
9									86.2–97.53	38	0.1674	0.52
10	≤40	≤0.5		≤40	≤ 0			≥23		37	0.1629	0.51
11									86.3–97.53	36	0.1585	0.51
12			≤0.7						86.92–106.35	35	0.1541	0.51
13			≤0.7						82.7–102.38	35	0.1541	0.51
14			≤0.7						86.92–104.36	34	0.1497	0.51
15			≥8.33		≤40				≥77.5	32	0.1409	0.51
16			≤0.7			≥20			≥77.9	31	0.1365	0.51
17			≤0.7				≤0		81.0–102.38	30	0.1321	0.51
18		≤0.5			≤0			≥23	≥79.0	28	0.1233	0.51
19			≤0.7						89.36–109.84	28	0.1233	0.51
20			≤0.7			≥20			81.0–121.77	26	0.1145	0.51

3.1. Predictive attribute significance

The results of the assessment of the significance of attributes in the decision rules are presented in Fig. 1. The higher the value of the condition attribute confirmation measure *s*, the greater its impact on the correct classification of objects. Attributes related to extrusion and spheronization, i.e., time and speed of spheronization, and number of die holes, have the greatest impact on the correct classification of objects. Among the excipients that significantly affect the classification of objects are amount of MCC, Genugel and lactose.

3.2. Results of stratified cross-validation

The average accuracy of the prediction is characterized by the parameters presented in Table 5. To validate the results obtained with VC-bagging, we performed the same type of stratified cross-validation with Random Forest and logistic regression implemented in WEKA toolkit (Frank et al., 2016). These results show that VC-bagging is producing at least as good results as Random Forest, which is considered as an off-the-shelf robust classifier allowing to obtain very good predictive accuracy. Both VC-bagging and Random Forest are producing better results than logistic regression.

4. Discussion

One of the basic methods of pellets production is a technique based on extrusion and spheronization. The main substance forming the core of the pellets is the MCC, which provides cohesiveness and binds the granules (Heng and Koo, 2001). Attempts to replace the MCC with other excipients encounter various technological problems at each stage of the pellets manufacturing process (Dukić-Ott et al., 2009). Applied substances have different physical and chemical properties and impact on water absorption by E/S mass or on properties of pellets (Dukić-Ott et al., 2009; Soh et al., 2008). These properties affect the quality of the extrudate obtained and its usefulness for spheronization. Application of overly plastic and sticky granules cause large, often oval pellets formation. Whereas too rigid and dry granules crumble rapidly into a fine fraction hardly subjected to spheronization (Umprayn et al., 1999).

The spherical granulate provides optimal pellet mass flow and easier coating (to obtain modified release of the active ingredient). Therefore, during the pellets production process spherical shape is one of the key parameters of their evaluation (Thommes and Kleinebudde, 2007; Soh et al., 2008; Umprayn et al., 1999; Podczeck et al., 1999; Koo and Heng, 2001). Initial evaluation of the shape of the pellets is a visual method. Disadvantage of this method is lack of representative sample and subjectivity of the investigator. However, it provides quick rejection of a series of poorly made pellets or possibility of E/S process modification. In order to conduct accurate and repetitive analysis of the shape microscopic techniques are combined with automatic evaluation of the shape of the examined image (Thommes and Kleinebudde, 2007; Soh et al., 2008; Umprayn et al., 1999; Podczeck et al., 1999; Koo and Heng, 2001). There are several ways to evaluate the shape of spherical particles in the literature. The most popular are: AR — aspect ratio, C — circularity, PS — projection sphericity, and e_R shape factor (Almeida-Prieto et al., 2007).

AR and e_R methods are most commonly used to evaluate spherical pellets. Circularity is not a recommended parameter for shape analysis because of the measurement errors in case of spherical irregularities (Thommes and Kleinebudde, 2007; Podczeck et al., 1999).

The most commonly studied factor influencing the shape of pellets is the amount of water in the extrudate (Thommes and Kleinebudde, 2007; Umprayn et al., 1999; Koo and Heng, 2001; Sousa et al., 2002; Lustig-Gustafsson et al., 1999; Podczeck and Newton, 2014; Fechner et al., 2003). Analysis of the decision rules received in class 1 (Table 3) points the presence of the “loss on drying” attribute in most rules. It is defined as a loss of water content from wet extrudate calculated in % in reference to dry mass. The most often appearing range is coinciding with the optimum observed by Thommes and Kleinebudde (2007), i.e. 81–117%. These are the amounts of water required to produce carrageenan pellets, which unlike MCC, binds more water during extrusion. Formulations containing MCC are described by decision rules with the attribute “loss on drying” ≤ 44 and ≤56%. The effect of water on the quality of the pellets is often described (Thommes and Kleinebudde, 2007; Soh et al., 2008; Podczeck et al., 1999; Almeida-Prieto et al., 2007; Sousa et al., 2002; Lustig-Gustafsson et al., 1999). The diversity of the water content required to obtain spherical pellets depends mainly on the various solubility of the active substances and excipients contained in the pellets (Lustig-Gustafsson et al., 1999).

Some problems arise with substances highly soluble in water. During extrusion, some substances dissolve in water to give solutions with increased viscosity (e.g. glucose). The addition of such substances to the pellet core results in a sticky extrudate. Extrudates sticking together make it difficult to spheronize and the resulting pellets are large. During drying the pellets hardness increase, which results in a higher density (Wlosnewski et al., 2010). The fact that parts of the substance dissolve in water cause a decrease in the amount of dry matter which must be wetted. The solution is to reduce the amount of water added during extrusion (Lustig-Gustafsson et al., 1999). The amount of water and the way it is bound to the pellet mass affects the extrusion process. The release of water while the mass is pressed through the sieve or extruder matrix reduces the shear strain that occurs during this process. In this case water acts as a lubricant (Wlosnewski et al., 2010). Increased shear strain during extrusion can cause a change in the structure or breakdown of MCC particles (Thommes and Kleinebudde, 2007; Fechner et al., 2003). This results in receiving a less rigid extrudate when carrageenan is applied. The “extrudates” plasticity has an impact on the form of the granulate during spheronization. The plasticity of the extrudate is necessary in order to acquire spherical shape of granules during spheronization (Thommes and Kleinebudde, 2007; Fechner et al., 2003). High water content in the extrusion mass cause problems during spheronization e.g. extrudates splicing. In result, the pellets are larger (Umprayn et al., 1999). Water or the liquid to solid ratio (L/S) has always an influence on pellet quality, also for poorly water-soluble substances (Thommes and Kleinebudde, 2007; Thommes and Kleinebudde, 2008; Lustig-Gustafsson et al., 1999). Most of the generated decision rules containing the API solubility attribute confirm the effect of low substance solubility in water (below 0.8 g/l). Some of the rules also define the content of calcium phosphate that is an insoluble excipient in the pellet mass.

Analyzed decision rules in a pellet class 1 specify the ranges of extrusion and spheronization parameters. They indicate higher spheronization rates (750 and 1000 rpm) as well as the number of die holes above 13. The condition attribute rank also shows that E/S parameters significantly affect classification (spheronization speed, spheronization time and number of die holes have the highest conformation coefficients *s*) (Fig. 1). If the number of die holes of extruder increase, the shear stress exerted on the mass extruded through the sieve is reduced. An important element of this process is the spheronization stage — process time and spheronizer rotation speed. During this process spherical pellets are formed. It was noted that the high spheronizer rotation (750 and 1000 rpm) compensates the hardness and rigidity of the extrudate (Thommes and Kleinebudde, 2007).

The energy delivered during such rotations is large enough for granules to become plasticized and the particles are given the right shape. Podczeck et al. noted that following parameters are important to compare or scale up the pellets production process: detailed parameters of the extruder and the spheronizer, the precision of the extrusion force, and the angular velocity at spheronization (Podczeck and Newton, 2014). These observations have been confirmed in the decision rules of the pellet set in class 2. The attribute “spheronization speed” was observed only in class 2, i.e. the rotation of the spheronizer < 750 rpm led to the incorrect pellets.

Table 4
Decision rules for class 2 pellets.

No.	API [%]	LogP	Solubility [g/l]	Lactose [%]	Mannitol [%]	Starch[%]	$CaHPO_4$ [%]	Gel911 [%]	MCC [%]	Screw speed [rpm]
1										
2			≤0.8							≤125
3			≤0.8							
4		≥1.2							≤0	
5										
6				≤0	≤0				≤0	
7					≤0				≤0	
8			≤0.8							≤125
9		≥1.2		≤0			≤0			
10		≥1.2					≤0		≤0	
11		≥1.2		≤0			≤0	≥20		
12	≥80								≤0	
13					≤0				≤0	
14		≥1.2								
15		≥1.2						≥20		
16										
17	≥80	≥1.2								≤100
18										
19	≥80								≤0	
20		≥0.5				≥40			≤0	

No.	number of die holes	Time of spheronization [s]	Spheronization speed [rpm]	Spheronization temperature [°C]	Loss on drying [%]	Rule support	Rule strength	Confirmation measure *s*
1	≤13		≤750.0		≤82.1	37	0.1629	0.82
2	≤13		≤750.0		≤82.1	33	0.1453	0.77
3	≤13		≤750.0	≤30.0	≤82.1	33	0.1453	0.71
4			≤750.0		≤82.1	21	0.0925	0.75
5	≤3		≤750.0		≤81.44	20	0.0881	0.69
6			≤750.0		≤73.5	18	0.0792	0.71
7			≤750.0		≤70.22	18	0.0792	0.69
8			≤750.0	≥30.0	≤81.0	16	0.0704	0.7
9			≤750.0	≤25.0		15	0.0660	0.7
10			≤750.0	≤25.0		15	0.0660	0.68
11			≤750.0	≤25.0		14	0.0616	0.73
12			≤750.0		≤75.3	13	0.0572	0.68
13			≤750.0		≤67.0	12	0.0528	0.7
14	≥23		≤750.0		≤82.1	12	0.0528	0.68
15	≥23		≤750.0		≤82.1	11	0.4845	0.68
16					74.3–77.3	10	0.0440	0.7
17			≤750.0			9	0.0396	0.7
18		≤240				9	0.0396	0.69
19			≤750.0		≤67.6	9	0.0396	0.68
20						7	0.0308	0.68

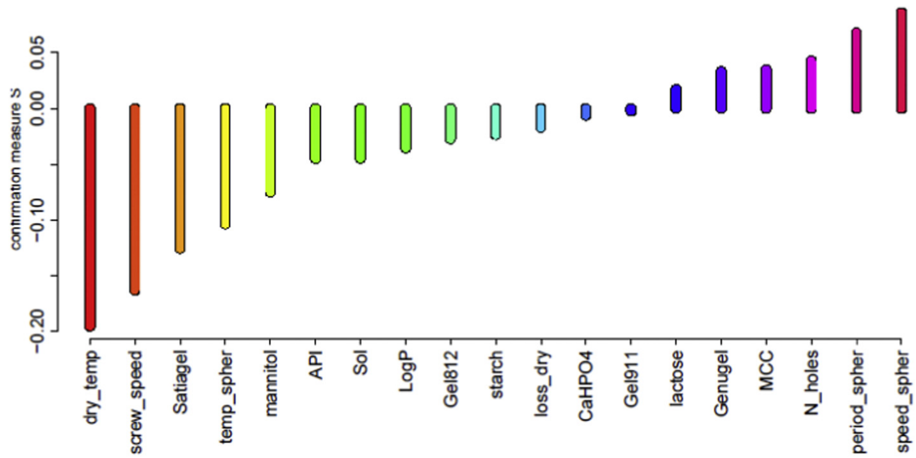

Fig. 1. Predictive attribute significance.

Table 5
Cross validation parameters.

	VC-bagging		Random forest		Logistic regression	
	[Avg. no.]	[Avg. %]	[Avg. no.]	[Avg. %]	[Avg. no.]	[Avg. %]
Correctly classified instances	184.12	81.13	182.14	80.24	174.76	76.99
Incorrectly classified instances	42.82	18.86	44.86	19.76	52.24	23.01
Average classification accuracy	78.47		77.60		73.98	
Average precision	81.25		80.13		76.56	

Mendyk et al. used artificial neural networks (ANNs) to analyze the data on which the pellet information system was based (Podczeck and Newton, 2014). ANNs proved their suitability for the pharmaceutical data analysis providing useful information about relationships governing pelletization procedure (Rojek et al., 2018; León Blanco et al., 2018). It was found that some pellet properties are more formulation-dependent, whereas others are more governed by API. Neural modelling allowed also to identify crucial variable sets for each of the analyzed problems based on the sensitivity analysis performed according to the described methodology. Achieved results were convergent to those presented in this paper. In the above-mentioned paper authors discretized certain ranges of technological parameters (high, medium, low) needed to obtain spherical or non-spherical pellets. The manner in which DRSA rules are presented using the actual range of parameters seems to be more reliable and easier to interpret.

According to Mendyk et al., ANNs were used as predictive models and data mining tools, thus leading to the identification of process critical variables and possibly in the future to the identification of some of the pelletization mechanisms (Mendyk et al., 2010). In our paper it is also shown which process parameters are of the greatest importance from the formulation point of view. Decision rules, based on DRSA, show ranges of technological properties and parameters to be included into pharmaceutical practice to obtain pellets with appropriate AR.

DRSA allows to discover relevant interactions between analyzed condition attributes. Obtained decision rules represent these interactions, showing conditions on selected attributes which jointly lead to a given result. In the analyzed information system the values of technological parameters (spheronization time, speed, and temperature, screw speed, and number of die holes) were almost the same in the whole system. It was the result of previous studies, as mentioned before. While these attributes do not enter to relevant interactions represented by decision rules, they cannot be omitted in the information system, because of their importance in the manufacturing of pellets.

Ronowicz et al. used the decision tree methodology (multivariate calibration technique) which allows to predict and to explain relationships between the preparation technology and the final product quality attribute (Ronowicz et al., 2015). A series of if-then rules provided deeper understanding and knowledge of factors affecting the pellet aspect ratio. The spheronization speed, spheronization time, number of holes and water content of extrudate have been recognized as the key factors influencing pellet aspect ratio. Our study revealed that the type and amount of excipients are also of great importance. In this way, rough set theory technique allows not only to plan the parameters of production but also it is able to predict the type and amount of excipients necessary to obtain pellets of required quality. Therefore, DRSA is more informative than decision tree methodology. Similarly, DRSA is not a black box method, because the decision rules are transparent and easy to interpret. Therefore, it can be even called a “glass box” (Greco et al., 2010), which is also in accordance with Quality by Design concept, because this approach implies better knowledge management, and more transparent decision making.

5. Conclusions

The analysis based on rough set theory allows to discover the most important relationships between the composition and the method of production on one side, and the quality of the investigated drug dosage forms on the other side.

The induced decision rules, along with Bayesian confirmation measures, allow to define the important parameters influencing the quality of the obtained drug dosage forms. To obtain spherical pellets the amount of water in the pellet mass, and the composition of the pellets (excipients used) were of the most important influence, taking into account considered technological parameters (spheronization time, speed, temperature, screw speed, and number of die holes).

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejps.2018.08.027.

Conflict of interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

Almeida-Prieto, S., Blanco-Méndez, J., Otero-Espinar, F.J., 2007. Microscopic image analysis techniques for the morphological characterization of pharmaceutical particles: influence of the software, and the factor algorithms used in the shape factor estimation. Eur. J. Pharm. Biopharm. 67, 766–776.

Błaszczyński, J., Greco, S., Słowiński, R., et al., 2009. Monotonic variable consistency rough set approaches. Int. J. Approx. Reason. 50, 979–999.

Błaszczyński, J., Słowiński, R., Stefanowski, J., 2010a. Variable consistency bagging ensembles. Trans. Rough Sets 5946, 40–52.

Błaszczyński, J., Słowiński, R., Stefanowski, J., 2010b. Ordinal classification with monotonicity constraints by variable consistency bagging. In: Rough Sets and Current Trends in Computing. Lecture Notes in Computer Science 6086. pp. 392–401.

Błaszczyński, J., Słowiński, R., Susmaga, R., 2011. Rule-based estimation of attribute relevance. In: Yao, J.T. (Ed.), RSKT 2011. LNCS 6954. Springer, Heidelberg, pp. 36–44.

Błaszczyński, J., Greco, S., Słowiński, R., 2012. Inductive discovery of laws using monotonic rules. Eng. Appl. Artif. Intell. 25, 284–294.

Bornhöft, M., Thommes, M., Kleinebudde, P., 2005. Preliminary assessment of carrageenan as excipient for extrusion/spheronisation. Eur. J. Pharm. Biopharm. 59, 127–131.

Chen, H.-L., Yang, B., Liu, J., Liu, D.-Y., 2011. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. Expert Syst. Appl. 38, 9014–9022.

Dukić-Ott, A., Thommes, M., Remon, J., Kleinebudde, P., Vervaet, C., 2009. Production of pellets via extrusion-spheronisation without the incorporation of microcrystalline cellulose: a critical review. Eur. J. Pharm. Biopharm. 71, 38–46.

Fechner, P.M., Wartewig, S., Füting, M., Heilmann, A., Neubert, R.H.H., Kleinebudde, P., 2003. Properties of microcrystalline cellulose and powder cellulose after extrusion/spheronization as studied by Fourier transform Raman spectroscopy and environmental scanning electron microscopy. AAPS PharmSci 5, 77–89.

Fekete, R., Zelkó, R., Marton, S., Rácz, I., 1998. Effect of the formulation parameters on the characteristics of pellets. Drug Dev. Ind. Pharm. 24, 1073–1076.

Frank, Eibe, Hall, Mark A., Witten, Ian H., 2016. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", fourth ed. Morgan Kaufmann.

Gardiner, E.J., Gillet, V.J., 2015. Perspectives on knowledge discovery algorithms recently introduced in chemoinformatics: rough set theory, association rule mining, emerging patterns, and formal concept analysis. J. Chem. Inf. Model. 55 (9), 1781–1803.

Greco, S., Matarazzo, B., Słowiński, R., 2001. Rough sets theory for multicriteria decision analysis. Eur. J. Oper. Res. 129, 1–47.

Greco, S., Pereira, R.A.M., Squillante, M., Yager, R.R., Kacprzyk, J. (Eds.), 2010. Preferences and Decisions: Models and Applications. Springer, Heidelberg.

Greco, S., Słowiński, R., Szczęch, I., 2016. Measures of rule interestingness in four perspectives of confirmation. Inf. Sci. 346–347, 216–235.

Heng, P.W.S., Koo, O.M.Y., 2001. A study of the effects of the physical characteristics of microcrystalline cellulose on performance in extrusion spheronization. Pharm. Res. 18, 480–487.

Koo, O.M.Y., Heng, P.W.S., 2001. The influence of microcrystalline cellulose grade on shape and shape distributions of pellets produced by extrusion-spheronization. Chem. Pharm. Bull. 49, 1383–1387.

León Blanco, J.M., González-R, P.L., Arroyo García, C.M., Cózar-Bernal, M.J., Calle Suárez, M., Canca Ortiz, D., Rabasco Álvarez, A.M., González Rodríguez, M.L., 2018. Artificial neural networks as alternative tool for minimizing error predictions in manufacturing ultradeformable nanoliposome formulations. Drug Dev. Ind. Pharm. 44 (1), 135–143.

Liu, H., Xiang, B., Qu, L., 2007. The application of rough sets in SAR analysis of N1-site substituted fluoroquinolones. Chemom. Intell. Lab. Syst. 87, 155–160.

Lourenço, V., Herdling, T., Reich, G., Menezes, J.C., Lochmann, D., 2011. Combining microwave resonance technology to multivariate data analysis as a novel PAT tool to improve process understanding in fluid bed granulation. Eur. J. Pharm. Biopharm. 78, 513–521.

Lustig-Gustafsson, C., Kaur Johal, H., Podczeck, F., Newton, J.M., 1999. The influence of water content and drug solubility on the formulation of pellets by extrusion and spheronisation. Eur. J. Pharm. Sci. 8, 147–152.

Margret Anouncia, S., Clara Madonna, L.J., Jeevitha, P., Nandhini, R.T., 2013. Design of a diabetic diagnosis system using rough sets. Cybern. Inf. Technol. 13, 124–139.

Mendyk, A., Kleinebudde, P., Thommes, M., Yoo, A., Szlęk, J., Jachowicz, R., 2010. Analysis of pellet properties with use of artificial neural networks. Eur. J. Pharm. Sci. 41, 421–429.

Pałkowski, Ł., Błaszczyński, J., Krysiński, J., Słowiński, R., Skrzypczak, A., Błaszczak, J., Gospodarek, E., Wróblewska, J., 2012. Application of rough set theory to prediction of antimicrobial activity of bis-quaternary ammonium chlorides. Lect. Notes Comput. Sci 7414, 107–116.

Pałkowski, Ł., Błaszczyński, J., Skrzypczak, A., Błaszczak, J., Kozakowska, K., Wróblewska, J., Kożuszko, S., Gospodarek, E., Krysiński, J., Słowiński, R., 2014. Antimicrobial activity and SAR study of new gemini imidazolium-based chlorides. Chem. Biol. Drug Des. 83, 278–288.

Pałkowski, Ł., Błaszczyński, J., Skrzypczak, A., et al., 2015. Prediction of antifungal activity of gemini imidazolium compounds. Biomed. Res. Int. 2015, 10 392326.

Podczeck, F., Newton, J.M., 2014. Influence of the standing time of the extrudate and speed of rotation of the spheroniser plate on the properties of pellets produced by extrusion and spheronization. Adv. Powder Technol. 25, 659–665.

Podczeck, F., Rahman, S.R., Newton, J.M., 1999. Evaluation of a standardised procedure to assess the shape of pellets using image analysis. Int. J. Pharm. 192, 123–138.

Rojek, B., Suchacz, B., Wesołowski, M., 2018. Artificial neural networks as a supporting tool for compatibility study based on thermogravimetric data. Thermochim. Acta 659, 222–231.

Ronowicz, J., Thommes, M., Kleinebudde, P., Krysiński, J., 2015. A data mining approach to optimize pellets manufacturing process based on a decision tree algorithm. Eur. J. Pharm. Sci. 73, 44–48.

Soh, J.L.P., Yang, L., Liew, C.V., Cui, F.D., Heng, P.W.S., 2008. Importance of small pores in microcrystalline cellulose for controlling water distribution during extrusion — spheronization. AAPS PharmSciTech 9, 972–981.

Sousa, J.J., Sousa, A., Podczeck, F., Newton, J.M., 2002. Factors influencing the physical characteristics of pellets obtained by extrusion-spheronization. Int. J. Pharm. 232, 91–106.

Srinivas, K., Rao, G. Raghavendra, Govardhan, A., 2014. Rough-fuzzy classifier: a system to predict the heart disease by blending two different set theories. Arab. J. Sci. Eng. 39, 2857–2868.

Swarbrick, J., 2007. Encyclopedia of Pharmaceutical Technology, third ed. Informa Healthcare, New York.

Thommes, M., Kleinebudde, P., 2006a. Use of κ-carrageenan as alternative pelletisation aid to microcrystalline cellulose in extrusion/spheronisation. II. Influence of drug and filler type. Eur. J. Pharm. Biopharm. 63, 68–75.

Thommes, M., Kleinebudde, P., 2006b. Use of κ-carrageenan as alternative pelletisation aid to microcrystalline cellulose in extrusion/spheronisation. I. Influence of type and fraction of filler. Eur. J. Pharm. Biopharm. 63, 59–67.

Thommes, M., Kleinebudde, P., 2007. Properties of pellets manufactured by wet extrusion/spheronization process using κ-carrageenan: effect of process parameters. AAPS PharmSciTech 8, 101–108.

Thommes, M., Kleinebudde, P., 2008. The behavior of different carrageenans in pelletization by extrusion/spheronization. Pharm. Dev. Technol. 13, 27–35.

Thommes, M., Blaschek, W., Kleinebudde, P., 2007. Effect of drying on extruded pellets based on κ-carrageenan. Eur. J. Pharm. Sci. 31, 112–118.

Tsumoto, S., 2000. Knowledge discovery in clinical databases and evaluation of discovered knowledge in outpatient clinic. Inf. Sci. 124, 125–137.

Uçar, T., Karahoca, A., Karahoca, D., 2013. Tuberculosis disease diagnosis by using adaptive neuro fuzzy inference system and rough sets. Neural Comput. & Applic. 23, 471–483.

Umprayn, K., Chitropas, P., Amarekajorn, S., 1999. Influence of process variables on physical properties. Drug Dev. Ind. Pharm. 25, 45–61.

Wilk, S., Słowiński, R., Michałowski, W., Greco, S., 2005. Supporting triage of children with abdominal pain in the emergency room. Eur. J. Oper. Res. 160, 696–709.

Wlosnewski, J.C., Kumpugdee-Vollrath, M., Sriamornsak, P., 2010. Effect of drying technique and disintegrant on physical properties and drug release behavior of microcrystalline cellulose-based pellets prepared by extrusion/spheronization. Chem. Eng. Res. Des. 88, 100–108.

Yu, L., 2008. Pharmaceutical quality by design: product and process development, understanding, and control. Pharm. Res. 25, 781–791.

Załuski, J., Szoszkiewicz, R., Krysiński, J., Stefanowski, J., 2004. Rough set theory and decision rules in data analysis of breast cancer patients. Lect. Notes Comput. Sci 3100, 375–391.