





Article

Evolutionary Algorithms in Modeling Aerodynamic Properties of Spray-Dried Microparticulate Systems

Adam Paclawski ¹, Jakub Szlęk ^{1,*}, Renata Jachowicz ¹, Stefano Giovagnoli ²,
Barbara Wiśniowska ³, Sebastian Polak ^{3,4}, Natalia Czub ¹ and Aleksander Mendyk ¹

¹ Faculty of Pharmacy, Jagiellonian University Medical College, Medyczna 9 Street, 30-688 Kraków, Poland; adam.paclawski@uj.edu.pl (A.P.); renata.jachowicz@uj.edu.pl (R.J.); natalia.czub@student.uj.edu.pl (N.C.); aleksander.mendyk@uj.edu.pl (A.M.)

² Department of Pharmaceutical Sciences, University of Perugia, Via del Liceo 1, 06123 Perugia, Italy; stefano.giovagnoli@unipg.it

³ Unit of Pharmacoepidemiology and Pharmacoconomics, Department of Social Pharmacy, Faculty of Pharmacy, Jagiellonian University Medical College, Medyczna 9 Street, 30-688 Kraków, Poland; b.wisniowska@uj.edu.pl (B.W.); sebastian.polak@uj.edu.pl (S.P.)

⁴ Certara UK–Simcyp Division, Blades Enterprise Centre, John Street, Sheffield S2 4SU, UK

* Correspondence: j.szlek@uj.edu.pl

Received: 7 August 2020; Accepted: 8 October 2020; Published: 13 October 2020



Abstract: Spray drying is a single step process in which solutions or dispersions are converted into dry particles. It is widely used in pharmaceutical technology to produce inhalable particles. Dry particle behavior during inhalation, described as the emitted dose (ED) and fine particle fraction (FPF), is determined in vitro by standardized procedures. A large number of factors influencing the spray drying process and particle interaction makes it difficult to predict the final product properties in advance. This work presents the development of predictive models based on experimental data obtained by aerodynamic assessment of respirable dry powders. Developed models were tested according to the 10-fold cross-validation procedure and yielded good predictive ability. Both models were characterized by normalized root-mean-square error (NRMSE) below 8.50% and coefficient of determination (R^2) above 0.90. Moreover, models were analyzed to establish a relationship between spray drying process parameters and the final product quality measures. Presented work describes the strategy of implementing the evolutionary algorithms in empirical model's development. Obtained models can be applied as an expert system during pharmaceutical formulation development. The models have the potential for product optimization and a knowledge extraction to improve final quality of the drug.

Keywords: spray drying; particles characteristic; evolutionary algorithms; empirical modeling; pharmaceutical technology; pulmonary drugs; microparticulate systems

1. Introduction

Current trends in pharmaceutical formulation development, reflected in guidelines for industry [1,2], indicate the need for a better understanding of manufacturing processes. The quality of pharmaceutical formulation depends on the qualitative and quantitative composition, applied production process, and its parameters. In-depth knowledge of production processes is crucial for quality assurance of the final product. Complexity, nonlinearity, and lack of full understanding of the processes are the main obstacles to the high-quality models development. Application of the artificial intelligence (AI) tools, such as evolutionary algorithms (EA), in the knowledge discovery of the manufacturing process, is beneficial for both the optimization of pharmaceutical formulations

and quality control purposes. The knowledge-based drug development is in accordance with process analytical technology (PAT) paradigm of drugs manufacturing advocated by the U.S. Food and Drug Administration agency (FDA). Moreover, the FDA pushes forward modeling as an important strategy for drug development and endorses terms “model informed drug development” (MIDD) and “model-based drug development” (MBDD). MIDD/MBDD extend substantially the classical PAT concept of drug manufacturing and navigates modern pharmaceutical industry towards Industry 4.0 and smart plants. Thus, it positions Information Communication Technologies (ICT) as pivotal elements of the successful development of new medicinal products. Since the pharmaceutical industry is highly regulated, to ensure this course the FDA announced that MIDD is going to be a subject of regulations and new guidelines for the industry in the future [3].

Spray drying is a single step process which converts a liquid feed into dry particles. It is widely used for producing inhalable microparticles including small molecules like salbutamol [4], ciprofloxacin hydrochloride [5] and macromolecules like cyclosporine A [6] or salmon calcitonin [7]. Spray-dried particles' properties depend on the type of equipment, working conditions, and feedstock characteristics [8]. Quality control procedures for pulmonary drug delivery systems imply the determination of the emitted dose (ED) and fine particle fraction (FPF) through the use of impactors and liquid impingers. Both parameters (ED and FPF) highly depend on assay conditions like apparatus, flow rate, inhaler type [9], and are determined in vitro by standardized procedures [10]. Multiple factors influence inhalable powder performances, such as particle shape, size, surface, charge, which determine inter-particle interaction [11].

Despite the wide use of the spray drying technology in the pharmaceutical industry, understanding of the process and the final product characteristics is still limited. Grasmeyer et al. [12] proposed an approach which allows prediction of outlet temperature and relative humidity of the exhausted air leaving the drying chamber. The first parameter was calculated based on heat loss caused by conduction and an evaporation process, while the second one was estimated by applying the Antoine equation and the ideal gas law. Simulations were consistent with experimental results. Furthermore, models were applied to optimize spray drying conditions for inulin and trehalose solutions. Another example is the application of computational fluid dynamics (CFD) to simulate the course of the spray drying process and prediction of outlet temperature, as well as residue accumulation during the operation [13]. The development of data-driven model presented by Amaro et al. focused on final product characteristics [14]. The statistical approach presented was successfully applied to predict particles size, and surface area for spray-dried trehalose and raffinose.

Soon enough, the pharmaceutical industry will be in the need for good quality, transparent (no black box allowed) predictive, and/or explanatory models, susceptible to validation performed at the highest level of pharmaceutical standards. Therefore, it is a strong need for the development of universal, transparent tools capable of predicting inhalable particles behavior. Regarding the complexity of the problem, we propose the application of evolutionary algorithms to develop high-quality models to predict FPF and ED based on controllable parameters of the spray drying process.

2. Materials and Methods

The results of laboratory experiments were used to construct the database. Dry powders were prepared by spray drying using Nano Spray Dryer B-90 (Büchi Labortechnik AG, Flawil, Switzerland). Feed solutions contained a model active pharmaceutical ingredient (API) and an excipient dissolved in the water–ethanol mixture. During spray drying, six factors were controlled and modified: API to excipient ratio [m/m%], feed solution concentration [m/V%], ethanol to water ratio [V/V%], inlet air temperature [°C], air flow [L/min], and pressure [mbar] inside the spray dryer. Particle aerodynamic performance was assessed by the Twin-stage Glass Impinger (Apparatus A, Copley Scientific, Colwick, UK) coupled with HandiHaler® inhaler. Assays were carried out according to the 10th Ph. Eur. monograph on aerodynamic assessment of fine particles for powder inhalers (2.9.18) [10]. The shape and size of obtained particles were analyzed based on Scanning Electron

Microscopy photographs using ImageJ software (1.52a, website: <https://imagej.net/>) [15]. The particles were spherical and their median size for a number distribution (Dn50) differed from 250 nm to 760 nm between formulations. The complete database, applied in further computational experiments, is shown in Table 1.

Table 1. Database applied in computational experiments.

API Content [m/m%]	Sol. Conc. [m/V%]	Ethanol [V/V%]	Inlet Air Temperature [°C]	Airflow [L/min]	Pressure [mbar]	FPF [%]	ED [%]
50	2	40	85	95	50	64.88	77.00
50	1	30	85	95	50	85.80	98.55
50	2	30	85	110	65	68.73	81.97
95	1	40	100	95	50	35.95	69.56
72.5	1.5	35	93	103	58	46.42	74.22
50	2	30	100	110	50	55.04	79.72
95	2	30	100	95	50	37.06	74.90
95	2	40	85	110	50	33.84	76.38
50	1	40	100	110	50	56.44	72.35
95	1	40	85	95	65	26.96	68.13
50	2	40	100	95	65	51.42	73.64
95	1	30	85	110	50	27.36	71.44
95	1	30	100	110	65	50.06	84.02
50	1	40	85	110	65	76.45	103.23
95	2	40	100	110	65	39.30	83.67
50	1	30	100	95	65	58.45	98.69
95	2	30	85	95	65	34.50	75.71

FPF—fine particle fraction; ED—emitted dose; API content—the mass ratio of API to excipient in feed solution; Ethanol—ethanol to water ratio in solvent applied in the process; Sol. conc.—the total concentration of substances dissolved in feed solution; Inlet air temperature—the temperature of the air entering spray drying chamber; Airflow—the amount of the air flowing by spray dryer chamber during one minute; Pressure—pressure inside spray dryer chamber.

2.1. Models Development and Assessment

The computational experiment was designed to develop multiple input single output (MISO) models. ED and FPF models were developed separately without any interaction of these two variables. Model performance was evaluated according to the 10-fold cross validation and expressed by three goodness of fit metrics: root-mean-square error (RMSE, Equation (1)), normalized root-mean-square error (NRMSE, Equation (2)), and coefficient of determination (R^2 , Equation (3)). NRMSE was calculated according to experimental FPF and ED ranges.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\text{pred}_i - \text{obs}_i)^2}{n}} \quad (1)$$

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

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

where: RMSE = root-mean-square error calculated for the model, X_{\max} = maximum value of the observed results, X_{\min} = minimum value of the observed results.

$$R^2 = 1 - \frac{\sum_{i=1}^n (\text{pred}_i - \text{obs}_i)^2}{\sum_{i=1}^n (\text{obs}_i - \text{obs}_{\text{mean}})^2} \quad (3)$$

where: R^2 = coefficient of determination, obs_i , $pred_i$ = observed and predicted values, respectively, obs_{mean} = arithmetic mean of observed values.

Models were built and assessed using in-house software developed in R environment [16]. Additional packages delivering evolutionary algorithms, fuzzy logic, and genetic programming methods were used. Computations were executed using 29 workstations (542 threads) working under Linux openSUSE Tumbleweed operating system. All tools applied in the study including source code and examples are available on SourceForge website [17–19].

2.2. Evolutionary Algorithms

Evolutionary algorithms gained research focus through the work of John Holland and his group in the 1970s, who introduced framework for genetic algorithm (GA) development [20]. Evolutionary algorithms are classified as heuristic search methods inspired by evolution process. The concept is based on the mechanism of natural selection where within the population only the fittest subjects, the ones which can adapt quickly to the solution, are selected for reproduction. The offspring inherits the characteristics of both parents, and if they better adapt to the observed solution (better fitness), they have a better chance to survive [21].

The given problem is defined by the environment, where a population of candidate solutions are present. Every candidate has a chromosome which is used to decode its information to the real solution. The computational space in the population is represented by the genotype where the phenotype represents its solution in the real world. In other words, chromosomes describe the proposed real solution of the problem, and cost function evaluates its goodness of fit. The evolution process usually starts by the random initialization of the population. Then, it iterates over three critical steps: (1) selection of the best-fitted individuals, (2) application of the genetic operators such as cross-over, mutation, and (3) reproduction—the next generation is created. These three steps are repeated until a termination condition is reached [22].

The process of model development and its selection was adapted from Polak et al. [23], where the authors used a similar methodology to explore the physiological parameters influencing QT interval, indicating the broad area of application of evolutionary algorithms. In this work, in-house software implementing two machine learning frameworks (*fugeR* and *rgp*) delivering computational methods based on evolutionary algorithms was employed.

2.2.1. *fugeR*

The first framework named *fugeR* [24] delivers solutions to develop fuzzy logic (FL) models by applying evolutionary algorithms fundamentals. Fuzzy logic was introduced by Zadeh [25], who claimed intermediate values between absolute true and absolute false, so called many-valued logic. The *fugeR* uses a genetic algorithm to evolve a population of the fuzzy logic models. In general, solution development starts by generating a random population of fuzzy systems, test them with available data, and sort them based on the goodness of fit metric. The following generations are created as decedents of the best performing models by applying cross-over function and random mutations. At the end of the process, the fuzzy system with the best performance is returned. The *fugeR* framework delivers enclosed functions to develop FL models using EA. However, to make model development process efficient and in line with the adopted standards, a wrapper software was created [18]. The system was built using Bash and R languages and works under Linux OS. It allows generating computational tasks with grid-like generated parameters including: learn and test dataset files, factors pass to *fugeR.run()* function for model development and scheme of models creation and assessment. Such approach allows testing developed models according to k-fold cross-validation standards and returns model performance as structured reports. Moreover, parallelization of the computational tasks speeds up the model development process and allows to find global problem solution. The general pipeline presenting the process of model development and testing process using the software is shown in Figure 1. During computational experiments, both model architectures and optimization settings

were tuned. The maximum number of rules differed from 10 to 50 with 1 to 20 variables within every fuzzy rule. The number of fuzzy membership functions (singletons) varied from 5 to 15. The learning strategy was controlled by population size (from 50 to 500) and generations number (from 100 to 500).

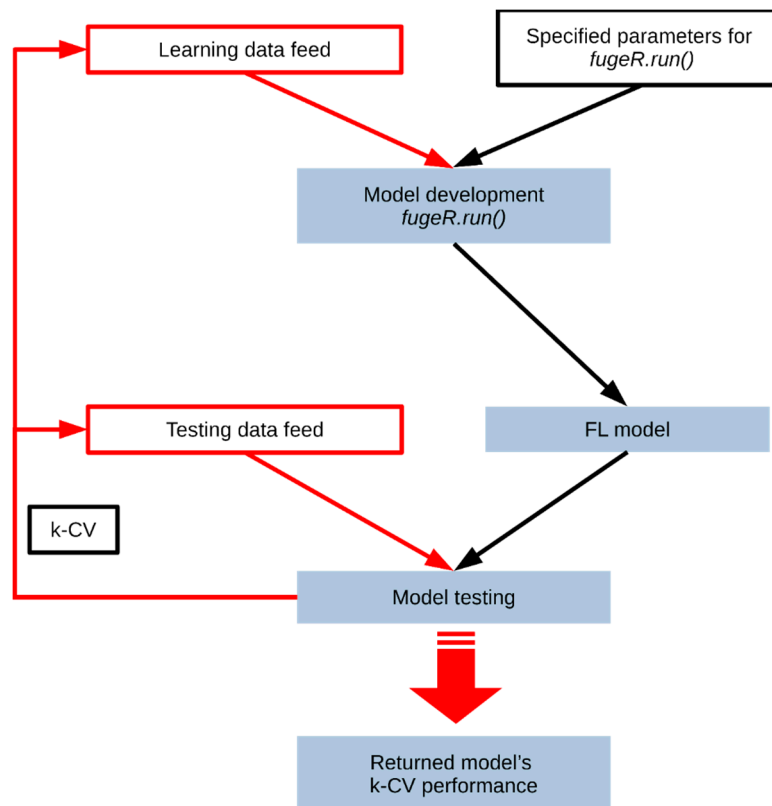


Figure 1. The workflow of in-house developed software implementing *fugeR* framework to model development and assessment system using fuzzy logic (FL) coupled with evolutionary algorithms. Software controls: data feed and model development process returning results in structured reports [18].

2.2.2. *rgp*

The second framework applied in presented work was *rgp* [26] developed by Olivier Flash, which delivers genetic programming methods. GP allows generating a computer program capable to solve a defined problem using EA. Moreover, using a symbolic regression module, the solution of a given problem can be represented as formal mathematical expression, which is easy to read and understand by humans. Similar to the previously described tool, at the first stage, the population of random equations is generated and tested. Next generations are created as descendants based on the best performing models by the exchange of a genetic material (cross-over function) and additional random modifications (mutations). To make the model development process efficient and in line with adopted methodology, in-house software for Linux OS was developed [19]. Source code was written in R and Bash programming languages. Software control parameters such as: data feed for the core function of model development through GP methods, flow of the evolution process (stop, resume), and performance metric the k-fold cross-validation (k-CV). Results are returned in structured report files. Model development and testing routine is continuous, which allows the user to follow recent advances and results. It is also possible to promptly use obtained models. The software is suited for parallelization of the processes by specifying a set of different parameters within the user interface. Moreover, computational tasks are run independently to speed up the model development process. The general scheme of the software workflow is shown in Figure 2. In the presented work, symbolic regression was used to develop a mathematical formula directly on the data provided. The size of the chromosome, representing the maximum length of the equation varied from 50 to 200. The population

size was set to 200, and the total length of the evolution process was set to 100 million steps divided into 500 testing stages. Moreover, the GP tools were used to perform feature selection based on the principle that variables related to the evolution of the solution are connected with the defined problem.

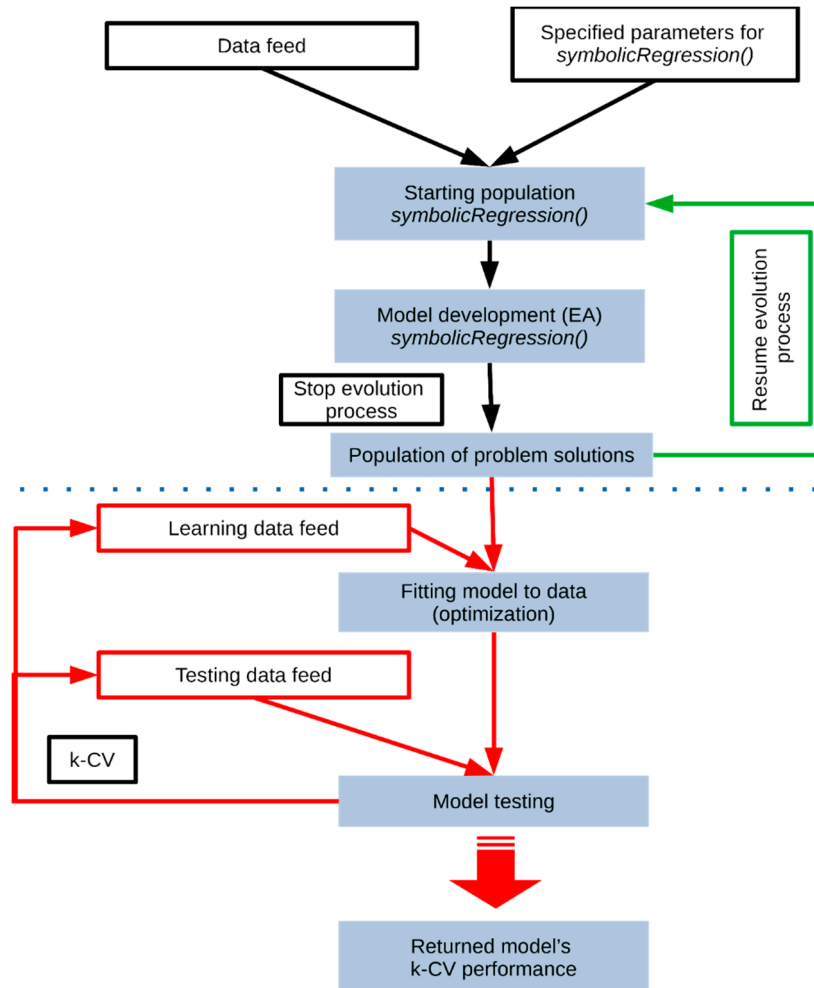


Figure 2. The workflow of in-house developed software implementing *rgp* framework to model development and assessment system using genetic programming methods. The first stage is based on a symbolic regression function allowing the development of formal mathematical equation fitting defined data. Later, the process is stopped, models are tested according to the k-fold cross-validation (k-CV) scheme, and the evolution process is resumed. Parameters affecting the evolutionary algorithms and model development process can be modified in the text user interface [19].

3. Results and Discussion

Prior to applying GP tools, linear regression models were developed and tested according to the 10-fold cross-validation scheme. Standard `lm()` function of R environment was applied, yet obtained results were far from being satisfactory. In the case of FPF prediction, RMSE was 12.03 (NRMSE = 20.44%), and in the case of ED predictions, RMSE was 11.30 (NRMSE = 32.2%).

3.1. Models for Prediction of FPF

The FPF prediction model performance is summarized in Table 2. Both *fugeR* and *rgp* models exhibited good FPF predictive ability. Fuzzy systems resulted in a slightly worse performance in comparison to genetic programming, as their NRMSE was 8.86 with R^2 below 0.85. The best model was developed using *rgp* package and it is presented as Equation (4) (NRMSE = 8.29%, and $R^2 = 0.91$).

The model uses four variables: API to excipient ratio (X_1), feedstock solution concentration (X_2), ethanol to water ratio (X_3), inlet air temperature (X_4), and five constants (C_1 – C_5). During the evolution of the solution, two variables describing flow rate and pressure inside the spray drying chamber were excluded. Comparison of FPF values predicted by the model and those obtained in laboratory experiments is presented in Figure 3.

Table 2. Summary of FPF prediction performance for models created with different methods.

	RMSE	R^2	NRMSE [%]
lm()	12.03	0.62	20.44
<i>fugeR</i>	5.21	0.84	8.86
<i>rgp</i>	4.88	0.91	8.29

FPF—fine particle fraction; RMSE—root-mean-square error; R^2 —coefficient of determination; NRMSE—normalized root-mean-square error.

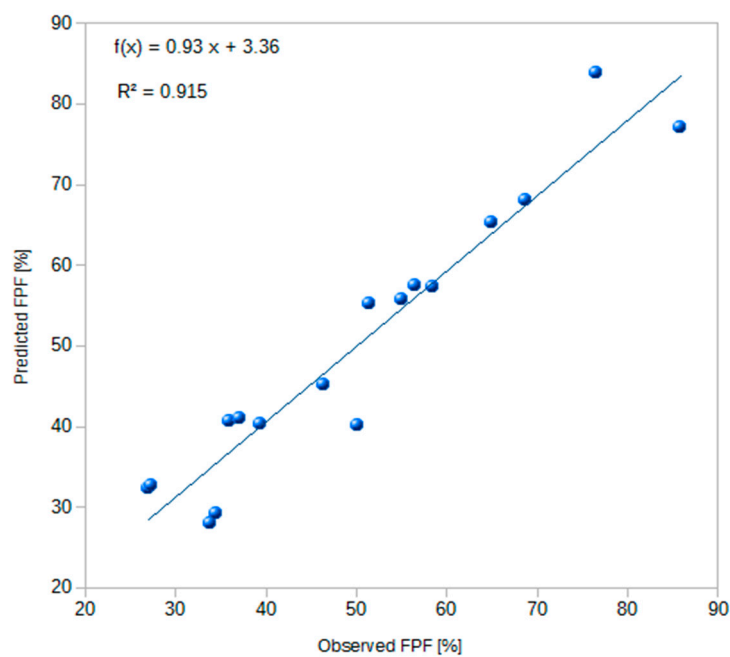


Figure 3. Comparison of FPF values predicted by the model and obtained in laboratory experiments. FPF—fine particle fraction.

$$FPF(\%) = \sqrt{[C_1 \cdot X_4 + X_2^2 \cdot \sqrt{(X_3)} + 2 \cdot X_1 + C_2]} \cdot \left[\sqrt{(2 \cdot X_4 + X_2^2 + C_3)} + C_4 \right] + C_5 \quad (4)$$

where:

C_1 – C_5 : constants; $C_1 = 5.298$; $C_2 = -524.223$; $C_3 = -135.864$; $C_4 = -13.842$; $C_5 = 120.661$,

X_1 : API to excipient ratio [m/m%],

X_2 : concentration of feed solution [m/V%],

X_3 : ethanol to water ratio in solvent applied in the process [V/V%],

X_4 : inlet air temperature during spray drying process [°C],

M: mass,

V: volume.

3.2. Models for Prediction of ED

Results obtained by the best fuzzy systems and GP models are presented in Table 3. Both methods resulted in a significantly lower error in comparison to linear regression. The GP models yielded almost twice-fold lower NRMSE (8.14%) than fuzzy systems (below 15%). The best model is given in Equation (5). ED values calculated according to Equation (5) and observed in laboratory experiments are shown in Figure 4. The model uses all six variables (X_1 – X_6) available in the database and two constants (C_1 and C_2).

Table 3. Summary of ED prediction performance for models created with different methods.

	RMSE	R ²	NRMSE [%]
lm()	11.30	0.15	32.20
fuzzyR	5.17	0.39	14.73
rgp	2.88	0.95	8.14

ED—emitted dose; RMSE—root-mean-square error; R²—coefficient of determination; NRMSE—normalized root-mean-square error.

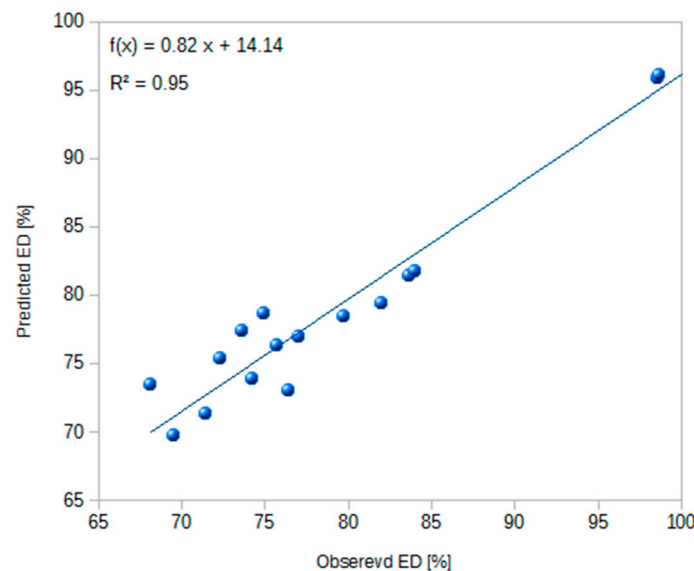


Figure 4. Comparison of ED values predicted by the model and obtained in laboratory experiments. ED—emitted dose.

$$ED(\%) = e^{e^{(\sin(X_6 + X_5 + \sin(\sqrt{X_4 - C_1})) + e^{(\sin(X_3)))})}} + e^{e^{(\sin(X_6 + X_2))}} + X_6 + e^{e^{(\sin(\sin(X_3 + X_2 - X_2)))}} + e^{e^{(\sin(X_2 + \sqrt{X_1}))}} + X_2 \quad (5)$$

- C_1 – C_2 : constants $C_1 = -5.029$; $C_2 = 8.417$,
- X_1 : API to excipient ratio [m/m%],
- X_2 : concentration of feed solution [m/V%],
- X_3 : ethanol to water ratio in solvent applied in the process [V/V%],
- X_4 : inlet air temperature during spray drying process [°C],
- X_5 : air flow during spray drying process [L/min],
- X_6 : pressure inside spray dryer during the process [mbar].

3.3. Model-Based Problem Analysis—Single Variable Impact

The predictive models were applied to explore the influence of input variables on predicted PPF and ED endpoints. Relationships between input and output variables are presented in Figure 5.

Analysis of Figure 5A shows that changing the ratio of API to excipient content strongly impacts ED and FPF values. Increasing the API content in the liquid feed decreases both the FPF and ED. The steepest descent of ED is observed when API content is in the range of 50 to 72.5% (Figure 5A). Figure 5B presents the influence of the total concentration of compounds dissolved in feed solution on the final particle characteristics. It is observed that lower concentrations of feed solution yield higher FPF and ED values. Ethanol to water ratio in solvent applied in the process has a small impact on FPF value while predicted ED is significantly lower for feed solvents with lower ethanol content (Figure 5C). The temperature of the air entering spray drying chamber does not affect ED at all (Figure 5D). On the contrary, low and high temperatures positively impact FPF values (Figure 5D).

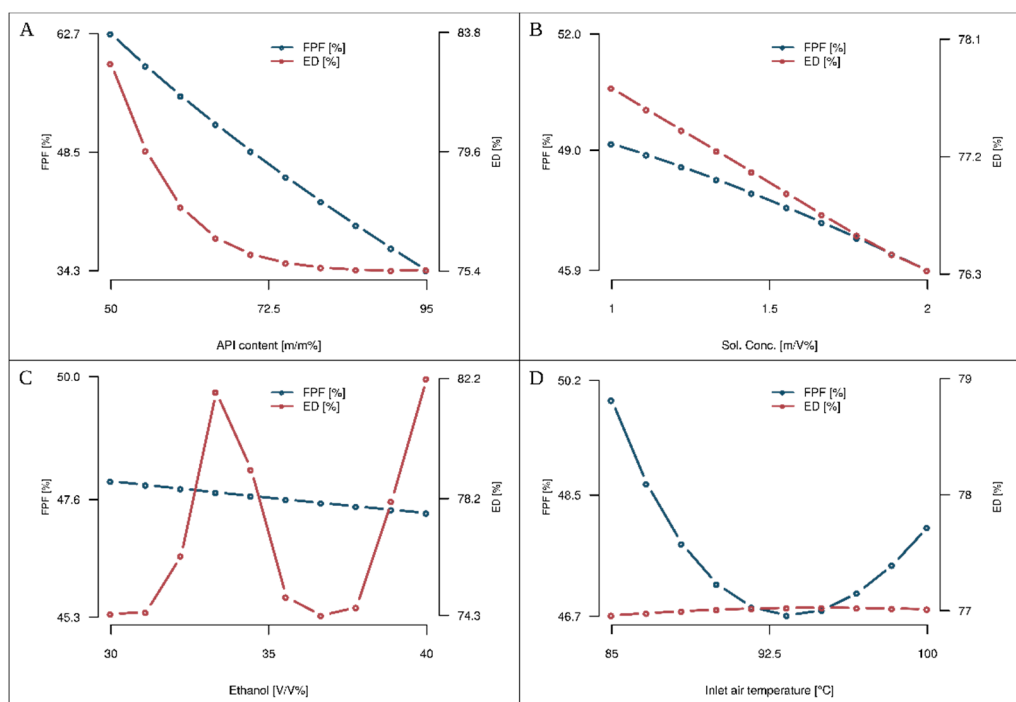


Figure 5. The relationship between ED and FPF predicted by models and numerical value of input variables: (A) API to excipient ratio in feed solution; (B) the total concentration of substances dissolved in feed solution; (C) ethanol to water ratio in solvent applied in the process; (D) the temperature of the air entering the spray drying chamber. FPF—fine particle fraction; ED—emitted dose.

3.4. Model-Based Problem Analysis—Multi-Variables Impact

Influence of the two input factors on the endpoints (FPF, ED) was analyzed by plotting three-dimensional graphs. Of a plethora of possible combinations, we have selected the most significant ones. Figure 6A presents the relationship between ethanol content in feed solvent, the total concentration of the solution, and predicted FPF value. It was observed that both low concentration of the feed solution and low ethanol content positively influence FPF values. Figure 6B shows the relationship between FPF predicted by the model, inlet air temperature during spray drying process, and API to excipient ratio. It can be noted that FPF value is higher for formulations containing lower API to excipient ratio. At the same time, inlet air temperature influences FPF in a non-linear manner. It appears that higher inlet air temperature is beneficial in case of formulations with high API excipient content while lower inlet air temperature results in higher FPF in case of formulations with lower API content. The influence of individual variables on the predicted ED value by the model is shown in Figure 7. It can be observed that decreasing API to excipient ratio positively influence predicted ED, and this trend is strong in the case of spray drying solution with a lower content of dissolved ingredients (Figure 7A). Figure 7B shows the influence of the inlet air temperature and solution concentration on

the predicted ED. Decreasing amount of feed solution concentration positively influences ED, whereas high and low inlet air temperatures seem to have a small positive impact on predicted ED.

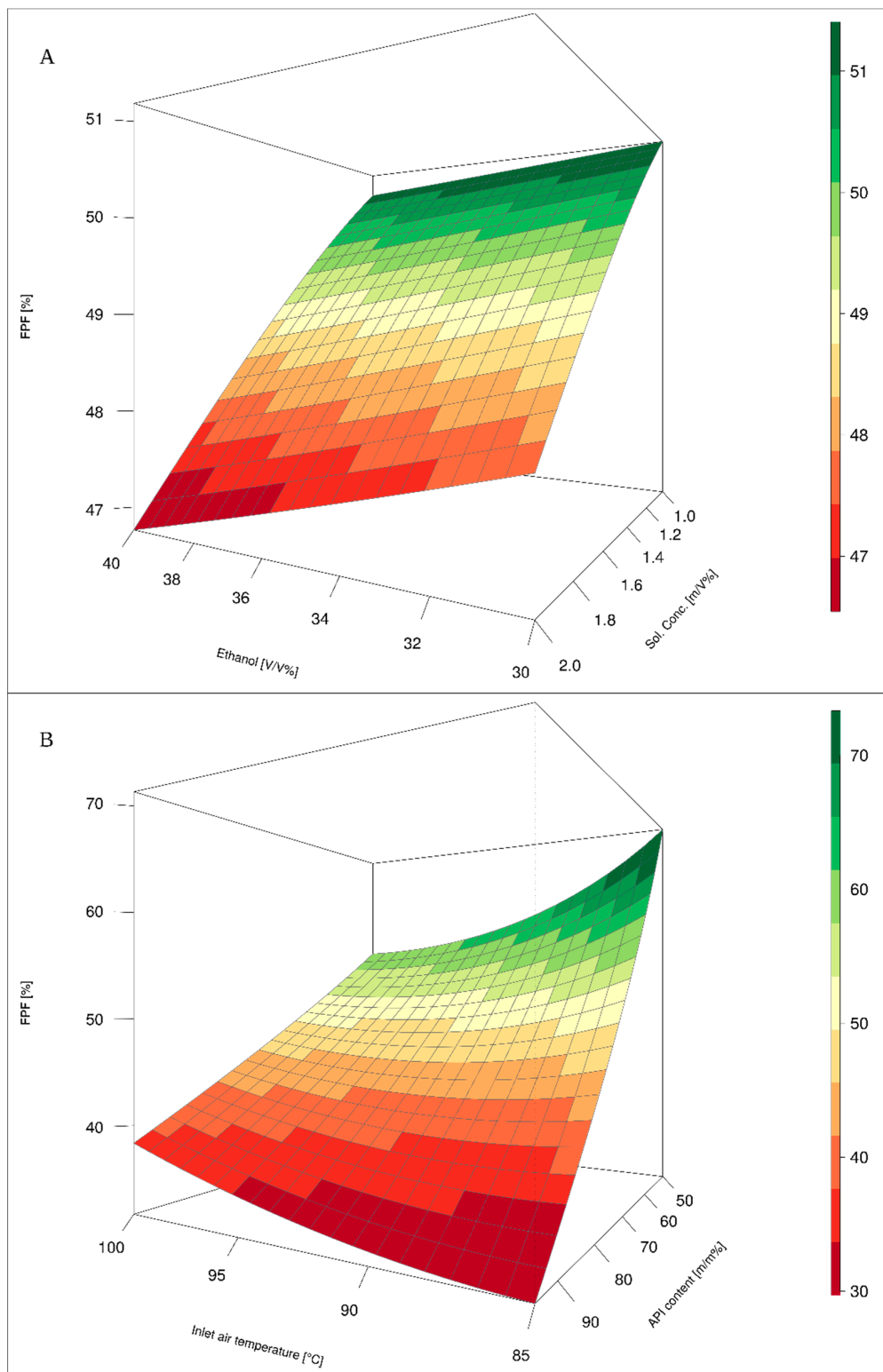


Figure 6. Factors influencing predicted FPF value: (A) ethanol to water ratio (Ethanol) and feed solution concentration (Sol. Conc.); (B) API to excipient ratio (API content) and inlet air temperature. FPF—fine particle fraction.

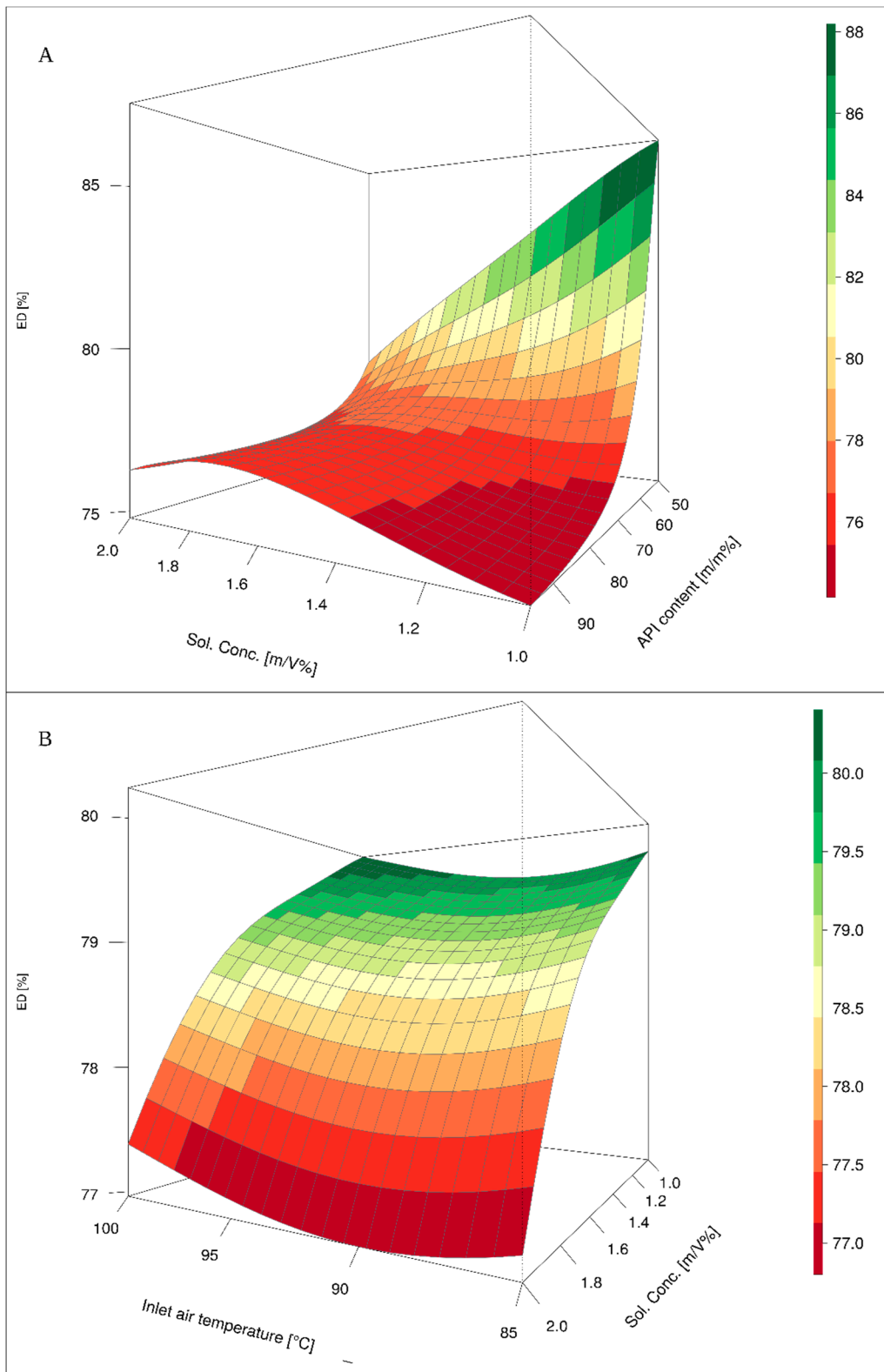


Figure 7. Factors influencing predicted ED value: **(A)** feed solution concentration (Sol. Conc.) and API to excipient ratio (API content), **(B)** inlet air temperature and feed solution concentration (Sol. Conc.). ED—emitted dose.

3.5. Discussion

The methodology applied during the development of evolutionary models stays in accordance with the Cross Industry Standard Process for Data Mining (CRISP-DM) protocol [27]. Therefore, the high quality of the models is assured. The obtained errors, below 8.3% for FPF and 8.2% for ED, in the 10-fold cross validation protocol suggests that the developed models exhibit significant extrapolation ability. With the latter confirmed, the models for ED and FPF could be used for prospective development of pulmonary drug delivery dosage forms, with models acting as decision support systems (DSS). This research demonstrates that the above presented models perfectly balance predictability with transparency [28,29]. Moreover, the GP derived models are transparent and easily validated (Equations (4) and (5)).

The model analysis suggests the highest impact of API to excipient ratio (Figure 5A) on the final aerosolization properties. These findings are in accordance with work of Bosquillion et al. [30], who observed higher ED values for spray-dried powders with lower API content. Moreover, positive impact on the FPF value with increasing amount of excipient was reported by Corrigan et al. [31] for spray-dried particles composed of salbutamol sulfate and lactose. It was also noted that decreasing total feed concentration resulted in the higher FPF. Chew et al. [32] observed influence of feedstock concentration on particles surface characteristic and powders aerodynamic properties. Low feedstock concentration led to more corrugated particles and resulted in the higher FPF, which corresponds with our analysis (Figures 5B and 6A). Ethanol content in the spray drying feed rate was also reported as a factor influencing particles morphology and ED [33], which confirms the validity of the selection of this variable by evolutionary algorithm (Equations (4) and (5), Figures 5C and 6A).

The limitations and the possibility to extend the presented model should also be pointed out. First of all, the discussed effects are valid for the substances applied in the experimental setup; therefore, in the case of other drugs and excipients it may differ in magnitude and direction. There is a possibility that increasing the other API concentration may result in higher ED or FPF experimental value, because feedstock and concentration effects are dependent on chemical–physical properties of the API, such as cohesiveness, hygroscopicity, and solubility. The same is with different spray drying setup including the type of apparatus, nozzle, and mechanism of aerosol generation, they may modify temperature impact on ED and FPF. Despite some limitations on the direct use of proposed models, it seems that the presented methodology could help in the development of optimal pharmaceutical formulation and understand underlying phenomena. Presented models could be improved and expanded in the future whereas methodology and workflow may remain unchanged.

4. Conclusions

Aerosolization properties of inhalable formulations are crucial for therapy effectiveness. Development of inhalation dry powders is currently based on a trial-and-error approach. The current study presents a way of developing models applying genetic algorithms, capable of predicting final powder aerosolization behavior based on initial spray drying conditions. Developed models were used for analysis of variable influence on the properties of the particles and delivered few insights about possible mechanisms governing selected endpoints. Moreover, the models can be applied as a predictive tool in terms of optimization or process control. Considering current trends in pharmaceutical technology reflected in the guidelines for industry, the presented approach can fulfill the requirements for better processes understanding and *in silico* design of the final product quality at an early development stage.

Author Contributions: Conceptualization, S.G., A.M., B.W. and S.P.; methodology, S.G., A.P., J.S.; software, A.M., A.P., J.S.; investigation, A.P.; resources, S.G., A.M.; data curation, A.P.; writing—original draft preparation, A.P.; writing—review and editing, S.G., A.M., J.S., N.C.; visualization, A.P., N.C.; supervision, A.M., R.J. and S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: A.P., A.M., and R.J. acknowledge Polish National Science Centre for the financial support (grant Symfonia 3 no 015/16/W/NZ7/00404). A.P. thanks Centre for Innovative Development of Society for grant allocation for mobility within ERASMUS + Programme.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. U.S. Department of Health and Human Services, Food and Drug Administration. *Guidance for Industry, PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*; CDER: Rockville, MD, USA, 2004.
2. U.S. Department of Health and Human Services, Food and Drug Administration. *Guidance for Industry, Q8(R2), Pharmaceutical Development*; CDER: Silver Spring, MD, USA, 2009.
3. Huang, S.M.; Zineh, I. Model-Informed Drug Development (MIDD): Opportunities and Challenges. Center for Drug Evaluation and Research US FDA, 2017. Available online: <https://www.fda.gov/media/104616/download> (accessed on 1 August 2020).
4. Littringer, E.M.; Zellnitz, S.; Hammernik, K.; Adamer, V.; Friedl, H.; Urbanetz, N.A. Spray Drying of Aqueous Salbutamol Sulfate Solutions Using the Nano Spray Dryer B-90—The Impact of Process Parameters on Particle Size. *Dry. Technol.* **2013**, *31*, 1346–1353. [[CrossRef](#)]
5. Karimi, K.; Pallagi, E.; Szabó-Révész, P.; Csóka, I.; Ambrus, R. Development of a microparticle-based dry powder inhalation formulation of ciprofloxacin hydrochloride applying the quality by design approach. *Drug Des. Dev. Ther.* **2016**, *10*, 3331–3343. [[CrossRef](#)] [[PubMed](#)]
6. Wu, X.; Zhang, W.; Hayes, D., Jr.; Mansour, H.M. Physicochemical characterization and aerosol dispersion performance of organic solution advanced spray-dried cyclosporine A multifunctional particles for dry powder inhalation aerosol delivery. *Int. J. Nanomed.* **2013**, *8*, 1269–1283. [[CrossRef](#)]
7. Kulvanich, P.; Sinsuepol, C.; Chatchawalsaisin, J. Preparation and in vivo absorption evaluation of spray dried powders containing salmon calcitonin loaded chitosan nanoparticles for pulmonary delivery. *Drug Des. Dev. Ther.* **2013**, *7*, 861–873. [[CrossRef](#)] [[PubMed](#)]
8. Maa, Y.F.; Costantino, H.R.; Nguyen, P.A.; Hsu, C.C. The effect of operating and formulation variables on the morphology of spray-dried protein particles. *Pharm. Dev. Technol.* **1997**, *2*, 213–223. [[CrossRef](#)] [[PubMed](#)]
9. Taki, M.; Marriott, C.; Zeng, X.-M.; Martin, G.P. Aerodynamic deposition of combination dry powder inhaler formulations in vitro: A comparison of three impactors. *Int. J. Pharm.* **2010**, *388*, 40–51. [[CrossRef](#)] [[PubMed](#)]
10. Council of Europe. Preparations for Inhalation: Aerodynamic Assessment of Fine Particles. In *European Pharmacopoeia*, 10th ed.; Council of Europe: Strasbourg, France, 2020; pp. 347–360.
11. Pilcer, G.; Amighi, K. Formulation strategy and use of excipients in pulmonary drug delivery. *Int. J. Pharm.* **2010**, *392*, 1–19. [[CrossRef](#)] [[PubMed](#)]
12. Grasmeijer, N.; De Waard, H.; Hinrichs, W.L.J.; Frijlink, H.W. A User-Friendly Model for Spray Drying to Aid Pharmaceutical Product Development. *PLoS ONE* **2013**, *8*, e74403. [[CrossRef](#)]
13. Goula, A.M.; Adamopoulos, K.G. Influence of Spray Drying Conditions on Residue Accumulation—Simulation Using CFD. *Dry. Technol.* **2004**, *22*, 1107–1128. [[CrossRef](#)]
14. Amaro, M.I.; Tajber, L.; Corrigan, O.I.; Healy, A.M. Optimisation of spray drying process conditions for sugar nanoporous microparticles (NPMPs) intended for inhalation. *Int. J. Pharm.* **2011**, *421*, 99–109. [[CrossRef](#)] [[PubMed](#)]
15. Mazzoli, A.; Favoni, O. Particle size, size distribution and morphological evaluation of airborne dust particles of diverse woods by Scanning Electron Microscopy and image processing program. *Powder Technol.* **2012**, *225*, 65–71. [[CrossRef](#)]
16. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018; Available online: <http://www.R-project.org/> (accessed on 1 August 2020).
17. Mendyk, A.; Szlęk, J.; Paclawski, A. R Scripts for Multivariate Analysis 2020. Available online: <https://sourceforge.net/projects/rscriptsmultivariate/files/> (accessed on 1 August 2020).
18. Mendyk, A.; Szlęk, J.; Paclawski, A. R Scripts for Multivariate Analysis: fugeR. 2020. Available online: <https://sourceforge.net/projects/rscriptsmultivariate/files/fugeR/> (accessed on 1 August 2020).
19. Mendyk, A.; Szlęk, J.; Paclawski, A. R Scripts for Multivariate Analysis: rgp. 2020. Available online: <https://sourceforge.net/projects/rscriptsmultivariate/files/rgp/> (accessed on 1 August 2020).

20. Holland, J.H. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, MI, USA, 1975.
21. Khan, M.W.; Alam, M. A Survey of Application: Genomics and Genetic Programming, a New Frontier. *Genomics* **2012**, *2*, 65–71.
22. Koza, J.R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*; MIT Press: Massachusetts, MA, USA, 1992.
23. Polak, S.; Wiśniowska, B.; Mendyk, A.; Paclawski, A.; Szlęk, J. Quantitative Assessment of the Physiological Parameters Influencing QT Interval Response to Medication: Application of Computational Intelligence Tools. *Comput. Math. Methods Med.* **2018**, *2018*, 1–11. [[CrossRef](#)] [[PubMed](#)]
24. Bujard, A. Fuger: Fuzzy Genetic, a Machine Learning Algorithm to Construct Prediction Model Based on Fuzzy Logic. R package version 0.1.2. 2012. Available online: <http://CRAN.R-project.org/package=fugeR> (accessed on 1 August 2020).
25. Zadeh, L.A. Fuzzy Sets. *Inf. Control.* **1965**, *3*, 338–353.
26. Flasch, O.; Mersmann, O.; Bartz-Beielstein, T.; Stork, J.; Zaefferer, J. rgp: R genetic programming framework. R. package version 0.4-1. 2014. Available online: <https://cran.r-project.org/src/contrib/Archive/rgp/> (accessed on 1 August 2020).
27. Shearer, C. The CRISP-DM Model: The New Blueprint for Data Mining. *J. Data Warehous.* **2000**, *5*, 13–22.
28. Neustein, A. *Text Mining of Web-Based Medical Content*; Walter de Gruyter: Berlin, Germany, 2014; 100p.
29. Paclawski, A.; Szlęk, J.; Nguyen, T.Q.N.; Lau, R.; Jachowicz, R.; Mendyk, A. Carrier optimization of pulmonary powder systems with using computational intelligence tools. *Powder Technol.* **2018**, *329*, 76–84. [[CrossRef](#)]
30. Bosquillon, C.; Lombry, C.; Pr eat, V.; Vanbever, R. Influence of formulation excipients and physical characteristics of inhalation dry powders on their aerosolization performance. *J. Control. Release* **2001**, *70*, 329–339. [[CrossRef](#)]
31. Corrigan, D.O.; Corrigan, O.I.; Healy, A. Physicochemical and in vitro deposition properties of salbutamol sulphate/ipratropium bromide and salbutamol sulphate/excipient spray dried mixtures for use in dry powder inhalers. *Int. J. Pharm.* **2006**, *322*, 22–30. [[CrossRef](#)] [[PubMed](#)]
32. Chew, N.Y.K.; Tang, P.; Chan, H.K.; Raper, J.A. How Much Particle Surface Corrugation Is Sufficient to Improve Aerosol Performance of Powders? *Pharm. Res.* **2005**, *22*, 148–152. [[CrossRef](#)] [[PubMed](#)]
33. Murtomaa, M.; Mellin, V.; Harjunen, P.; Lankinen, T.; Laine, E.; Lehto, V.-P. Effect of particle morphology on the triboelectrification in dry powder inhalers. *Int. J. Pharm.* **2004**, *282*, 107–114. [[CrossRef](#)] [[PubMed](#)]



  2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).