*Research Article*

# Integrating Artificial and Human Intelligence into Tablet Production Process

Matjaž Gams,[1] Matej Horvat,[2] Matej Ožek,[1] Mitja Luštrek,[1] and Anton Gradišek[1,3]

***Abstract.*** We developed a new machine learning-based method in order to facilitate the manufacturing processes of pharmaceutical products, such as tablets, in accordance with the Process Analytical Technology (PAT) and Quality by Design (QbD) initiatives. Our approach combines the data, available from prior production runs, with machine learning algorithms that are assisted by a human operator with expert knowledge of the production process. The process parameters encompass those that relate to the attributes of the precursor raw materials and those that relate to the manufacturing process itself. During manufacturing, our method allows production operator to inspect the impacts of various settings of process parameters within their proven acceptable range with the purpose of choosing the most promising values in advance of the actual batch manufacture. The interaction between the human operator and the artificial intelligence system provides improved performance and quality. We successfully implemented the method on data provided by a pharmaceutical company for a particular product, a tablet, under development. We tested the accuracy of the method in comparison with some other machine learning approaches. The method is especially suitable for analyzing manufacturing processes characterized by a limited amount of data.

**KEY WORDS:** artificial intelligence; machine learning; process analytical technology; process optimization; tablet manufacture.

## INTRODUCTION

Manufacturing of pharmaceutical products from raw materials includes a wide variety of parameters—all of which have to be precisely adjusted in advance in order to ensure good quality of the final product. Parameters range from the precursor raw materials attributes, such as particle size distribution, moisture content, crystallinity levels, or other physical properties, to the production process parameters, such as temperatures, pressures, and holding times at different stages of the manufacturing process. While fine tuning of this multitude of parameters is crucial for the optimized quality of the final product, manual tuning is not an efficient approach as it is time-consuming and has to be repeated every time the attributes of the new batch of raw material change.

In 2004, the United States Food and Drug Administration (FDA) issued a document "PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance" (1). This document was written as guidance for the pharmaceutical industry with a broad audience in different organizational units and scientific disciplines. The goal of Process Analytical Technology (PAT) initiative is to understand and control the manufacturing process flexibly in real time—which will provide pharmaceutical industry flexible systems for designing, analyzing, and controlling manufacturing processes (2). The FDA expects an inverse relationship between the level of process understanding and the risk of producing a poor quality product. PAT-inspired quality control is based on in-process electronic data rather than laboratory testing on a subset of the final product (as in the classical approach); therefore, essentially, the entire production run may be evaluated for quality control purposes and may require less restrictive regulatory approaches to manage change. The new solution enables similar results as the existing approaches, but enables greater flexibility, i.e., adapting to a wider range of the input options before and during the process. In addition, it enables greater flexibility inside the target, e.g., adapting to a specific subgroup, a subtarget inside the target, and a potential future application. The comparison between a classical and a PAT-inspired production process is shown in Fig. 1.

Similar to PAT, Quality by Design (QbD) initiative is intended to be a modern substitute to the classical pharmaceutical quality-by-testing paradigm in which the product quality is ensured by raw material testing, a fixed drug product manufacturing process, in-process material testing, and end-product testing. QbD is aiming to achieve superior quality with as little testing as possible by focusing the testing on a few critical parameters and attributes that affect the product

[1] Department of Intelligent Systems, Jožef Stefan Institute, Jamova 39, SI-1000, Ljubljana, Slovenia.
[2] Sandoz Biopharmaceuticals Mengeš, Kolodvorska 27, SI-1234, Mengeš, Slovenia.
[3] To whom correspondence should be addressed. (e-mail: anton.gradisek@ijs.si)

**ABBREVIATIONS:** AI, Artificial intelligence; ML, Machine learning; PAT, Process Analytical Technology; QbD, Quality by Design.
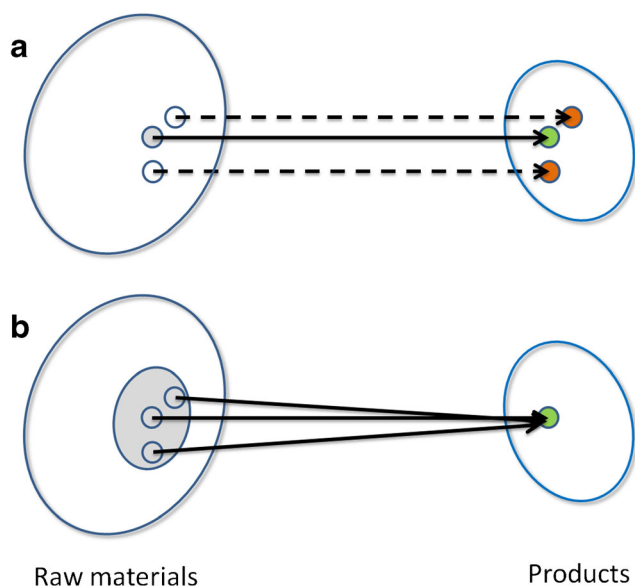
**Fig. 1.** Comparison between the classical and the PAT-inspired production process. In the classical production process, **a** the path from a very narrow raw material subspace (*gray circle*) to a good-quality product (*green circle*) is strictly defined. A fixed production path from slightly different starting material (*white circles*) results in rejected products (*red circles*). Using PAT, **b** the flexibility of the process path allows producing good quality product from a much wider raw materials subspace

and the process the most. QbD elements may include the following (3):

– Design and development of product and manufacturing processes
– Definition of quality target product profile
– Identification of critical quality attributes, critical process parameters, critical material attributes, and sources of variability
– Control of manufacturing processes to produce consistent quality over time

Within QbD, process and product understanding is ideally achieved during design and development phase. Nevertheless, data from manufacturing experience is a proven, valuable, and historically most frequent source of knowledge. Often during development, the scope of process and raw material variability is highly limited, thus preventing full exploration of the operating space. On the other hand, the scope of observed variability increases throughout the product life cycle. Therefore, any method that can learn from production data is beneficial and complementary to the concept of QbD.

Many researchers are trying to implement concepts introduced by PAT and QbD to the production process. This can be achieved in various ways, here divided into three groups:

– Introduction of sensors and automatic logging system into the production line
– Introduction of new methodologies into the process design
– Improving existing processes by data analysis

Regarding the sensors, for example (4), presents a review of a variety of experimental techniques used to monitor the hot-melt extrusion process.

When trying to implement the PAT and QbD guidelines, researchers often use projection-based methods (e.g.,

PCA—Principal Component Analysis, PLS—Partial Least Squares) for data analysis. These are statistical methods that transform a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components (5,6). For example, such approach was applied to crystallization process through several case studies (6).

There are several machine learning (ML) algorithms for manufacturing process control (overview can be found in (7)). H. Sadoyan *et al.* (8) used a combination of kNN clustering and rule extraction (on an industrial example of rapid tool making) while Cotofrei and Stoffel (9) performed ML on events in order to construct a set of rules with various induction techniques.

Several techniques exist for analyzing large datasets. They are suitable for processes where data is gathered automatically in short time intervals. However, in the case of pharmaceuticals manufacture, the amount of data may be limited both in time resolution, of measurements (lack of automated process data logging at the production site), as well as in the number of available examples due to the high cost of production or low number of produced batches due to market demand considerations. Moreover, in contrast to a continuous process, where the relationship between the inputs and the outputs is continuously monitored, the batch quality can only be confirmed at the end of the run.

Here, we introduce a novel technique for improving existing pharmaceutical production processes by interaction between human and artificial intelligence (AI), e.g., between AI programs and a human operator. The method uses the data obtained from the initial production batches and is optimized to work on a relatively small amount of initial data, due to high costs of obtaining it in the test runs. The data is used to construct decision trees which are then transformed into human-understandable rules that will guide the production process. Supplementary guidance is provided by automated machine learning and statistical methods, which are not understandable (intuitive), but offer a global viewpoint as another opinion compared to the local-oriented rules. The results are presented to the human operator through specialized visualization techniques, which allow operators to test several possible future settings of parameters and inspect the resulting AI predictions. Our method was developed in cooperation with a pharmaceutical company, using the data of a particular product, a tablet, under development. Our approach differs from the related work in the variety of the methods used, combined together with human operators into one sequential iterative procedure.

## PROBLEM DESCRIPTION

We address the problem of controlling an immediate release tablet manufacturing process using process and quality data gathered in the past. The tablet manufacturing process for our particular product consists of six unit operations

– Raw material dispensing
– High-shear wet granulation
– Vacuum drying and cooling
– Sieving
– Final blending and lubrication
– Tablet compression

Finally, the tablet samples are tested and the quality of the end product is established.

From a regulatory point of view, only two quality classes matter—accepted and rejected. In reality, the quality is a continuum, and any number of quality classes can be defined. Data from manufacture indicated that in this particular product, there is a tendency towards a high variability of tablet assay between the defined sampling locations during the compression process, with possible linkage to an increased variability of blend uniformity results and occasional segregation tendencies at the beginning and the end of the compression process. Therefore, the quality class was determined by jointly considering the following quality testing results:

– Final blend uniformity results
– Assay variability within sampling location
– Assay variability between different sampling locations
– Deviation of assay from target at the beginning and the end of compression

For our purpose, based on the above inputs, judgments of three independent human experts were employed to assign each batch to one of the following classes of product quality: (1) high quality, (2) medium quality, and (3) rejected. The judgments were predominantly based on the four quality testing results listed above. Each expert considered the result values and concluded on a final classification. Additionally to the quality results, supplementary batch information data, such as IPC values, batch record comments, etc., were considered by the experts.

The data on which the guided supervision of the manufacturing process was to be based was collected from 29 batches of product (29 examples in ML terminology) of 4 different tablet strengths. The set of batches included process development and validation batches. The four different tablet strengths are proportional in mass and composition, and are produced by the same technological process (apart from tablet compression). Our goal was to exploit the inherent variability present within the available development, validation, and production data in order to extract additional valuable knowledge. From the available data, each batch was initially characterized by 71 features (27 raw material attributes, marked as S1–S27, and 44 process parameters, marked as P1–P44). With the help of ML methods and expert guidance (as discussed in the following section), we have managed to decrease the number of features down to the 54, shown in Table I. Nevertheless, manual pruning of the features was done cautiously and only those were removed for which some thorough prior evidence of non-relevance was available.

## MATERIALS AND METHODS

The approach for improving pharmaceutical production consists of two phases: in the first phase, the classifiers to predict the quality of the product are constructed, and in the second phase, these classifiers are used interactively by a human operator to guide the production.

The first phase begins with grouping the features for machine learning based on the stage of the production. The term *features* refers both to the attributes of the raw materials and the parameters of the production process. During the batch production, data is gathered gradually by process stages, and decisions are taken at each stage; therefore, the values of the features of the remaining stages of the process are not known until the end of the current stage. To account for this fact, the features are grouped into training sets corresponding to the process stages: each set contains the features of its stage and all the preceding stages. The stages and their features are presented in Table I: the first stage is dispensing of raw materials (batch initial conditions); the second stage is the first step of the production process, etc., until the last stage, the compression of the tablets. Using several smaller subsets of the data enables 5–10% higher classification accuracy than creating classification rules based on a complete, non-segmented dataset.

After the grouping step, we proceed with the two main steps of the first phase, which are motivated by the two goals we had. First, we wanted to present the main influences on each prediction about the quality of the product, to help the human operator to make informed decisions. Second, we wanted to maximize the accuracy of the prediction. The first goal was achieved by using rules generated from decision trees, which the operator can understand. This step is described in the "Rules Construction" subsection. The second goal was achieved by combining the rules with non-transparent ML methods, which is described in the "Meta-learning" subsection.

The meta-learning step completes the first phase, which is followed by the second phase described in the "Human-Computer Interaction" subsection. All the steps of our approach are shown in Fig. 2.

## Rules Construction

The rules to predict the quality of the tablets were constructed in four steps. In the first step, decision trees were built. Decision trees on small datasets such as ours tend to overfit the data, meaning that they describe the training data too precisely and do not generalize well to new data. Because of that, in the second step, only the parts of the trees found best and most reliable by human experts were kept, while the rest of the trees were discarded. In the third step, rules were built from the best parts of the decision trees. In the fourth step, the rules were combined (also taking care that the combination of rules was non-conflict) into a single classifier.

Decision trees are used as the basis for the rules construction because they are one of the most popular forms of knowledge representation. Their key advantage is that they are understandable even to non-experts on ML. Among disadvantages, their classification accuracy is often not as good as that of less understandable methods. For the construction of decision trees, the C4.5 algorithm (10) implemented in Weka (11) ML suite was used. The trees were constructed from each of the seven training sets corresponding to the seven stages of the production. Building several decision trees instead of only one offers two important advantages. First, they increase the reliability of the classification, as observed in several well-known ML methods that rely on multiple classifiers, such as bagging, boosting, and ensemble methods (12,13). Second, more decision trees contain more features. This is important for covering the complete design space (14), since experts believe that every feature that they find important should be included in the classifier.

**Table I.** Selected Raw Material Attributes and Process Parameters

| Name | Stage | Features (attributes and parameters) |
|---|---|---|
| S01–S21 | Raw materials | Active pharmaceutical ingredient: particle size distribution (Malvern) Microcrystalline cellulose: loss on drying, sieve analysis (through 0.075 and 0.215 mm), conductivity Lactose: sieve analysis (through 0.250 and 0.100 mm), bulk volume, tapped volume, loss on drying |
| P01 | Tablet strength | Tablet strength (four strengths, proportional tablet design) |
| P02–P15 | High-shear wet granulation | Granulation parameters (two predefined parameter settings), nozzle spray pattern (narrow or wide), initial amount of granulation fluid, impeller torque after initial addition of granulation fluid, product temperature after initial addition of granulation fluid, additional amount of granulation fluid, impeller torque after additional amount of granulation fluid, impeller torque difference, final product temperature, difference between initial and final product temperature, total wet-massing time |
| P16-P25 | Vacuum drying | Vacuum pressure, product temperature after initial drying phase, sieving mesh size, vacuum pressure of final drying phase, final product temperature, loss on drying, possible additional drying cycle with same parameters as the final drying phase, total drying time |
| P26–P29 | Cooling | Vacuum pressure, product temperature, cooling time, yield |
| P30–P31 | Sieving | Mesh size, yield |
| P32–P44 | Tablet compression | Tableting speed, filling speed, main pressure, filling depth, normalized filling depth (dose-dependent), minimum punch gap, normalized minimum punch gap, de-dusting pressure, tablet press ID (multiple presses of same make and model), average tablet thickness, normalized average tablet thickness, tablet thickness variability, hardness, hardness variability, disintegration time (minimum, maximum). |
| Class | | Batch quality class (1 = high, 2 = medium, 3 = rejected) |

At the end of the tree construction, we had a set of trees that mostly overfit the data. To improve the accuracy on new data as well as the understandability, rules were extracted from the trees by human experts. They were assisted by tools that find intersections and unions of the features used by individual rules, and check for redundancy. We also made sure that every important feature was used at least once. The resulting collection of rules was termed the *Tablet classifier*.

**Meta-learning**

While the objective of the rules construction step was to understand the main influences of features on each prediction about the quality of the product, we also wanted to maximize the accuracy of the predictions. This is typically achieved by testing several algorithms and selecting the most accurate one

(11,12,15). We used C4.5 decision trees, support vector machines, bagging on decision trees, AdaBoost on decision trees, and the *Tablet classifier* from the previous subsection. The standard ML algorithms were implemented in the Weka ML suite.

Since a combination of several algorithms typically outperforms any single algorithm, we also tested this approach. We combined the same algorithms that were compared individually. First, they were combined by majority voting, which means that the final class was the one predicted by most of the classifiers constructed by the individual ML algorithms. Second, they were combined by meta-learning, which means that a meta-classifier was constructed to combine the single classifiers. The meta-classifier was trained with the C4.5 ML algorithm on the same features as the single classifiers (subsets of features from Table I), to which the predictions of the single classifiers were appended.
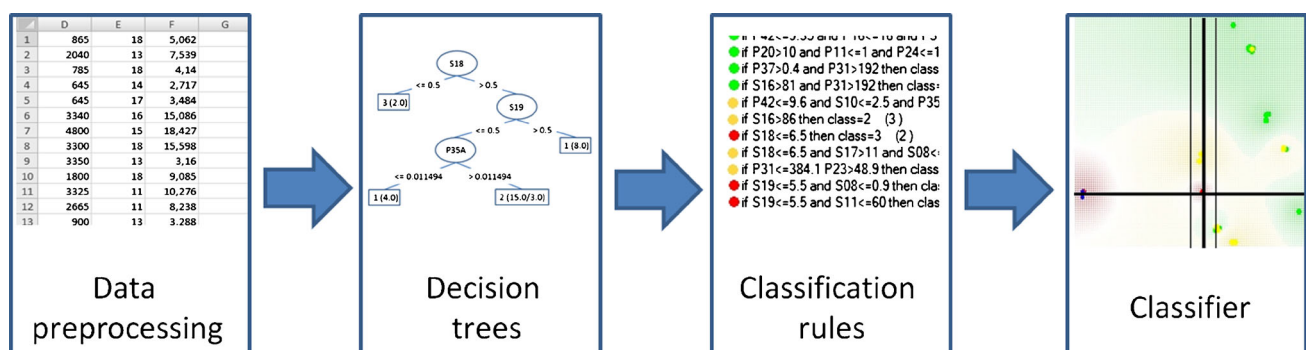


**Fig. 2.** Schema of the ML process. Several decision trees are automatically built, rules are extracted by experts, and meta-classifiers are used to produce the final classifier
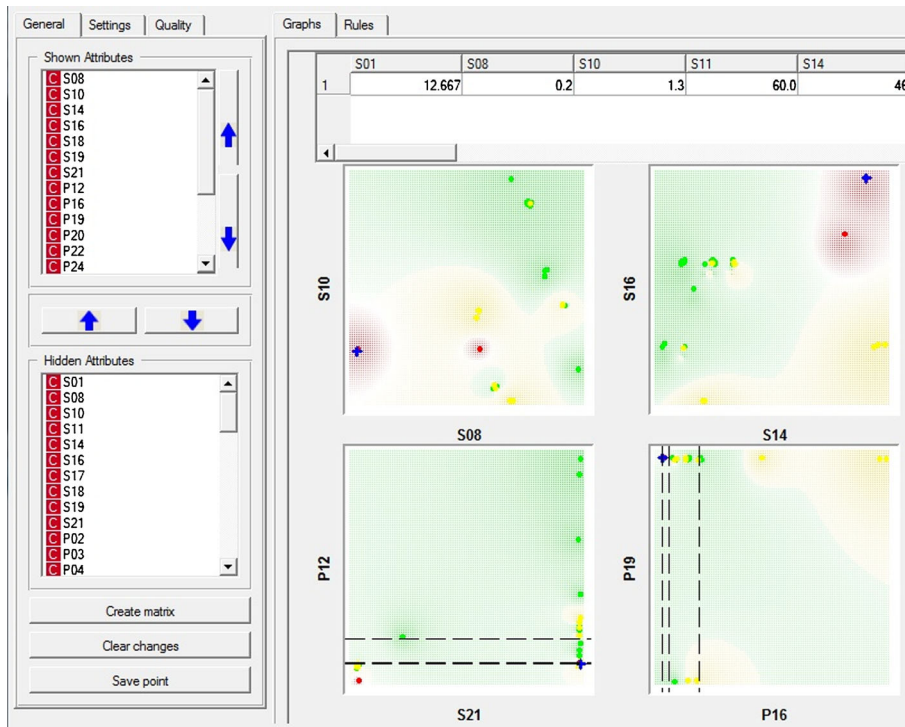
**Fig. 3.** Interface of the visualization, modeling, and testing procedure

## Human-Computer Interaction

The *Tablet classifier* and meta-classifier are used as tools for the human operator to control the manufacturing process. There are many parameter settings that the operator would need to examine at each step to successfully guide the process. The operator needs not only the predicted product quality as a consequence of a particular parameter setting but also an understanding of the reasons for the prediction. To achieve such understanding, the most relevant rules are extracted from the rule set and presented to the operator. The quality predicted is presented as a single decimal number, computed with the meta-classifier as described in the previous subsection (1–3, the product class).
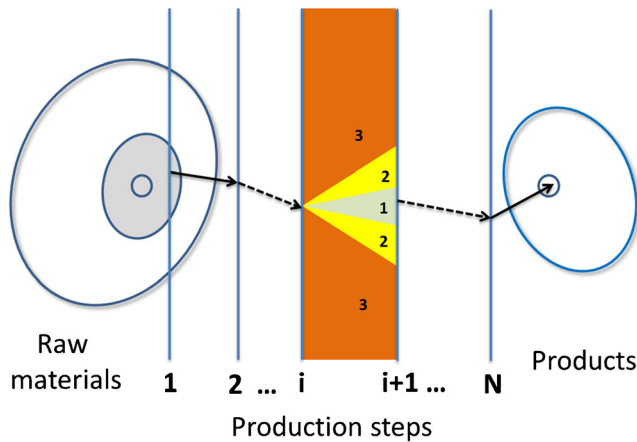


**Fig. 4.** Our approach based on the PAT initiative. The precursor raw materials are outside the narrowly defined subspace for classical production. Our method guides the expert user to choose the optimal parameters at each step in the process. At the *i*th step, the method indicates the intervals where the parameters, if chosen, will lead to high (*1*) or medium (*2*) quality product, or the product will be rejected (*3*)

To facilitate the decisions, the graphic user interface presents to the operator a set of two-dimensional diagrams that are projections of the N-dimensional space (Fig. 3). At each step, only the most relevant projections are shown to the operator, since not all of them are important at the same time. The most important attributes and parameters are those whose changes will have the largest impact on the product quality. This has been taken into account previously, in the construction of the *Tablet classifier*. The operator can interactively investigate a potential change in the selected parameter values in order to check the improvement of the product quality. Parameters are changed by clicking on the diagrams. At the same time, the *Tablet classifier* predicts the quality of the end product. If the change causes some other features to become problematic, they will automatically be shown in a separate window with explanation provided. Since a specifically trained and knowledgeable operator has expert knowledge of the production process, he is able to make reasonable decisions at each step. Therefore, our method will assist him to guide the process to a good-quality final product. A schematic representation of this approach is show in Fig. 4. Here, it should be stressed that within a regulated environment,

**Table II.** Comparison of the Classification Accuracy of Different Machine Learning Methods

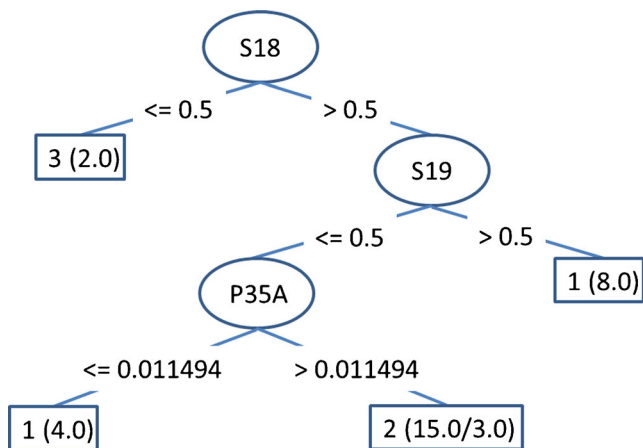| Method | Classification accuracy (%) |
|---|---|
| Decision tree | 72 |
| Support vector machine | 90 |
| Bagging | 85 |
| AdaBoost | 93 |
| Tablet classifier | 91 |
| Majority voting | 97 |
| Meta-classifier | 99.7 |

**Fig. 5.** Decision tree. *S18* and *S19* are raw material attributes; *P35A* is a process parameter

there are some limitations to the nature and extent of changes that are allowed. Therefore, additional care must be taken during the implementation of the system to ensure adequate qualification of involved personnel, equipment, and overall quality systems.

### RESULTS

To assess the quality of our *Tablet classifier*, we compared its results to the results obtained using the four other classifiers listed in the "Meta-learning" subsection. Since the learning examples were sparse in the problem space, methods like cross-validation or leave-one-out achieve accuracies close to 100%. To reasonably compare the methods, we generated synthetic test examples by introducing small modifications in the descriptions of existing examples (e.g., we slightly changed the values of some features in the real data to generate a much

bigger training dataset). The accuracies of final predictions of methods are presented in Table II.

A single decision tree has the lowest classification accuracy. Decision trees alone performed significantly worse than the human-modified rules (*Tablet classifier*). The use of human expert knowledge in design of the rules therefore provides a significant advantage over these automated methods. The only single classifier that outperformed the *Tablet classifier* was constructed by the AdaBoost algorithm, since this algorithm builds many decision trees and combines them. It achieved 21 percentage points higher accuracy than a single tree and 2 percentage points higher than the *Tablet classifier*. AdaBoost has the usual advantage of machine learning—being able to find subtle relations in the data and quantitatively evaluate them, and additionally combines multiple trees, covering more features than single trees (feature coverage is important [16]).

Figure 5 shows a single decision tree built on data for the particular tablet. The C4.5 algorithm places the most important attribute on the top of the tree; the second most important attribute, in relation to the top attribute, is the second from the top, and so on. Apparently, the parameter S18 (lactose bulk volume in ml/5 g) is the key parameter; if smaller than 0.5, the product will potentially be rejected (denoted by class 3 in the left leaf of the tree). If S18 is larger than 0.5, then parameter S19 should be tested (lactose tapped volume in ml/5 g). If S19 is larger than 0.5, the product will be of high quality (class 1). Otherwise, we continue down the tree, until we finish in one of the leaves. The numbers in parentheses denote the numbers of examples in the training data belonging to the leaf, and the number after the slash denotes the number of examples with the parameters belonging to the leaf, but with a different class (i.e., misclassified examples).

Figure 6 shows a meta-classification decision tree, which also includes the outputs of single classifiers as features. One can see that the most important parameters are the outputs of AdaBoost and *Tablet classifier*, and parameters at the end of the process.
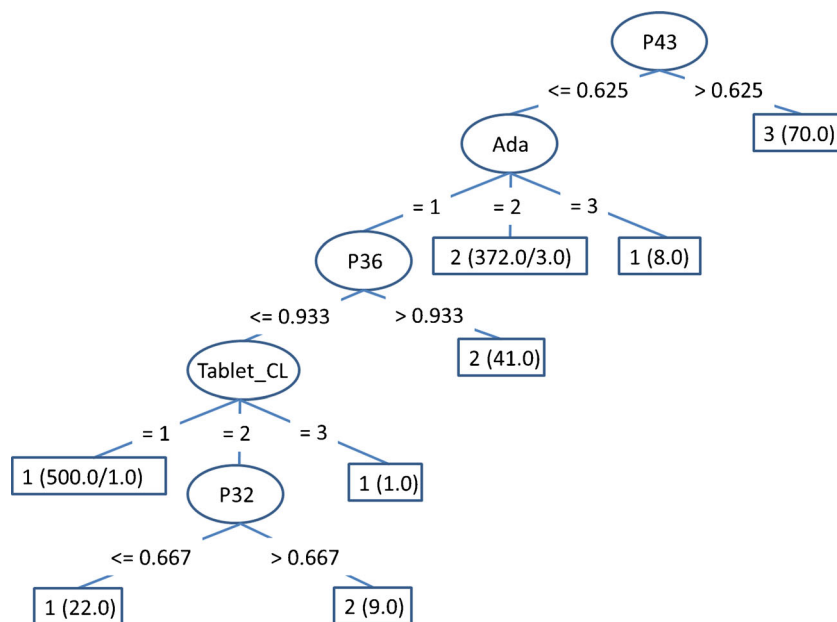


**Fig. 6.** The meta-classification decision tree. *Ada* and *Tablet_CL* are the outputs of AdaBoost and *Tablet classifier*

## DISCUSSION AND CONCLUSION

In this paper, we present a new method for the analysis and guidance of tablet production process, using the concepts of the PAT and QbD initiatives. We developed a new ML method specialized for process analysis on small databases. It combines rule construction with the help of human experts and the induction of an ensemble of classifiers. To facilitate the interaction with the operator, we created an application module for the visualization and guidance of the manufacturing process. The method is not intended to be used autonomously, but as an interactive aid to a human operator. It enables the operator to simulate small changes in the manufacturing process—with the system providing the visualization of the consequences.

This can be illustrated by the following example. An explorer wants to cross a frozen lake but does not know whether the ice is thick enough to hold his weight. This is essentially a three-dimensional problem where the parameters are the coordinates on the lake surface ($x$ and $y$) and the ice thickness $z$. The program simulates the next step (in the $xy$ direction), checks whether the thickness of ice is sufficient, and produces the outcome (*yes/no*). In case of a negative outcome, the explorer can check the $xz$ view to see whether the ice is too thin at that spot. After choosing the best step, the explorer proceeds with the next round of simulations and eventually crosses the lake safely. The manufacturing process, on the other hand, contains a significantly higher number of dimensions (features)—from raw material attributes to process parameter settings. At each step, the operator needs to review many settings to successfully guide the process, and our method guides him through the procedure.

Our method presents, highlights, and explains only the most important data and events to avoid information overload. Additional information can be found on additional screens when one needs supplementary explanation for the classification. We believe that the chosen representation with several two-dimensional graphs enables an important insight into the ongoing production process and enables effective human-computer interaction. Humans test, overview, compare, and decide, while computers simulate consequences of events, provide explanation, and enable quantitative and qualitative comparisons of various decisions.

The presented method has some limitations. First of all, it is specialized for the task with a limited data, and might not work as accurately as the existing methods on more typical amounts of data. Moreover, the significant amount of expert tuning of the method in the first stages prevents automatic updating of the method when new data is available. Further research into how such an approach could be automated is desired.

## ACKNOWLEDGMENTS

## REFERENCES

1. Guidance for Industry: PAT—a framework for innovative pharmaceutical development, manufacturing, and quality assurance. U.S. Department of Health and Human Services, Food and Drug Administration, 2004.
2. Schneidir R. Achieving process understanding—the foundation of strategic PAT programme. Processing in Pharmaceutical manufacturing, 2006.
3. Yu LX. Pharmaceutical quality by design: product and process development. Pharm Res. 2008;25(4):781–91.
4. Saerens L, Vervaet C, Remon JP, De Beer T. Process monitoring and visualization solutions for hot-melt extrusion: a review. J Pharm Pharmacol. 2014;66:180–203.
5. Kenett RS, Kenett DA. Quality by design applications in biosimilar pharmaceutical products. Accred Qual Assur. 2008;13(12):681–90.
6. Yu LX, Lionberger RA, Raw AS, D'Costa R, Wu H, Hussain AS. Applications of process analytical technology to crystallization process. Adv Drug Deliv Rev. 2004;56(4):349–69.
7. Wang XZ. Data mining and knowledge discovery for process monitoring and control. London: Springer; 1999.
8. Sadoyan H, Zakarian A, Mohanty P. Data mining algorithm for manufacturing process control. Int J Adv Manuf Technol. 2006;28:342–50.
9. Cotofrei P, Stoffel K. Rule extraction from time series databases using classification trees. Proceedings of the 20th IASTED Conference on Applied Informatics, Innsbruck, 2002.
10. Quinlan JR. C4.5: Programs for machine learning. Morgan Kaufmann; 1993.
11. Witten IH, Frank E. Data mining: practical machine learning tools and techniques. 3rd ed. San Francisco: Morgan Kaufmann; 2011.
12. Gams M. Weak intelligence: through the principle and paradox of multiple knowledge. New York: Nova Science Publishers, Inc; 2001.
13. Maudes J, Rodríguez JJ, García-Osorio C, Pardo C. Random projections for linear SVM ensembles. Appl Intell. 2011;34(3):347–59.
14. Mishra A, Bhatwadekar N, Mahajan P, Karode P, Banerjee S. Process Analytical Technology (PAT): boon to pharmaceutical industry. Pharm Rev. 2008; 6(6).
15. Russell SJ, Norvig P. Artificial intelligence: a modern approach. 3rd ed. Upper Saddle River: Prentice Hall; 2010.