

# 機械学習モデルのスクリーニングによる製剤製造因子の抽出

## Extraction of Pharmaceutical Production Factors by Screening of Machine Learning models

酒井憲一<sup>\*1</sup> 吉村志保<sup>\*1</sup> 山村尚弘<sup>\*1</sup> 太田智明<sup>\*1</sup> 山中祐治<sup>\*1</sup> 古賀明子<sup>\*1</sup>  
Kenichi Sakai Shiho Yoshimura Takahiro Yamamura Tomoaki Ohta Yuji Yamanaka Akiko Koga

<sup>\*1</sup> 中外製薬株式会社 生産工学研究部  
Chugai Pharmaceutical Co., Ltd. Production Engineering Department

Improving manufacturing conditions for stable manufacturing is important from the viewpoint of stable supply of pharmaceutical products. The manufacturing process of pharmaceutical products is strictly controlled under GMP, but the quality varies to some extent during continuing the production. If this variance can be reduced, more stable manufacturing becomes possible. The purpose of this study was to extract latent manufacturing factors that lead to a reduction of variance of dissolution ratio, one of the quality parameters of a pharmaceutical product, using machine learning. To derive machine learning models, DataRobot, an automated machine learning platform, was used. By the use of multiple models with high prediction accuracy, which were selected by screening of models, we evaluated the influence of manufacturing factors comprehensively from various viewpoints. As a result, "granulation-water temperature" could be extracted as a latent factor. By lowering this temperature, it was estimated that the variance of dissolution ratio reduces.

## 1. はじめに

適正な品質を有する医薬品を安定的に製造し、患者さんに提供し続けることは、製薬メーカの重要なミッションの一つである。適切な品質を担保した医薬品を安定的に製造するため、製造工程は適正製造規範(GMP)の下で厳しく管理されている。このような状況においても、製造を重ねていくなかで、製品品質はある程度の範囲で変動する。この変動を抑えることができれば、製造の頑健性が高まり、より安定的に生産が可能となるが、そのための製造因子を見つけ出すことは容易ではない。

これまで、Fig. 1 に示したプロセスで製造したカプセル剤のデータを用いて、医薬品有効成分の溶け具合を示す“溶出率”と呼ばれる品質を目的変数とした単回帰や重回帰分析を試みてきたが、溶出率の変動要因を見つけることはできなかった。そこで、本検討では、データに潜むパターンを発見しモデル化する機械学習の手法を適用することにした。機械学習モデルの導出には、自動機械学習プラットフォームである DataRobot を用いた。モデルをスクリーニングし、予測精度の良い複数のモデルを用いることで、製造因子の影響を多様な視点から総合的に評価し、溶出率に影響を及ぼす潜在的因子の抽出を行った。

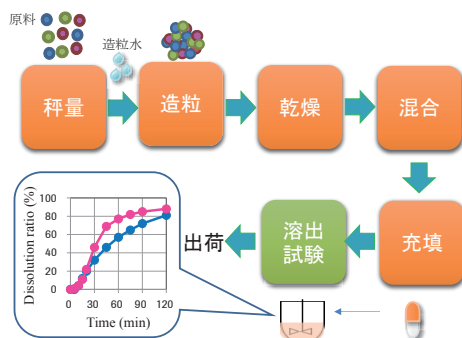


Figure 1. カプセル製造フロー。原料に水を加え造粒、乾燥したものをカプセル充填した後、溶出試験で溶出率を評価。

## 2. 方法

### 2.1 教師データ

Fig. 1 に示したプロセスで製造したカプセル剤 240 ロットを対象に機械学習モデルを構築した。

目的変数は、各ロットの溶出試験で得られた溶出率( $n = 6$ で溶出試験を実施し、その中の最低値)とした。カプセル剤の有効成分が完全に溶解すると、溶出率は 100%を示す。

説明変数(特徴量)としては、以下に示す 6 つのプロセスパラメータを用いた。

- 造粒工程:
  - ①スプレー速度(kg/min)、
  - ②造粒水温度(°C)、
  - ③造粒消費電力(kW)
- 乾燥工程:
  - ④乾燥時間(min)、
  - ⑤乾燥終点時の排気湿度(%)
  - ⑥乾燥終点時の排気温度(°C)

### 2.2 機械学習モデルのスクリーニング

自動機械学習プラットフォームである DataRobot(ver 4.3.1、DataRobot, Inc., MA)を用いて、数十個の学習器でモデルを導出し、10 分割交差検証を行って、それぞれのモデルを平均二乗誤差平方根(RMSE)で評価した。アンサンブルモデルを除いた上で、RMSEを指標にモデルをスクリーニングし、予測精度がよい複数のモデルを用いることで、製造因子の影響を、多様な視点から総合的に評価できるようにした。

### 2.3 潜在的製造因子の抽出

Permutation Importance [Breiman, 2001]により求めた特徴量インパクト(Pi)を DataRobot から得た後、選択したモデルにおける説明変数の影響度を式(1)により求めた。Pi は、モデルの中で最もインパクトが大きい変数を特徴量インパクト 100%とし、他の変数はインパクトに応じて相対的な値を与えられている。

影響度と平均絶対誤差率 (MAPE) との比較、ドメイン知識に基づいて、製法の改良に資する製造因子を抽出した。  
製法改良を想定し、抽出した製造因子の管理値を変更した場合の溶出率の改善度を推定した。

$$Impact \% = \frac{P_i}{\sum_i P_i} \times 100 \quad (1)$$

3. 結果と考察

得られたモデル群から、RMSE が低い(予測精度が高い)上位 3 つのモデル (Model 1~3) を Table 1 に示した (Model 1~3)。用いられた学習器は決定木 (RandomForest、ExtraTree) と SVM であった。それぞれの MAPE は 10% 弱であった。

Table 1. RMSE 上位モデル

Model No.	Model name	Cross-Validation (RMSE)	MAPE (%)
Model 1	RandomForest Regressor	5.5732	8.9
Model 2	Nystroem Kernel SVM Regressor	5.5795	8.8
Model 3	ExtraTree Regressor	5.5967	8.9

各モデルにおける説明変数の影響度と、モデル間での平均値を Table 2 に示した。造粒消費電力、乾燥終点時の排気湿度、乾燥終点時の排気湿度、造粒水温度、乾燥時間、スプレー速度において、平均影響度はそれぞれ 37.7、19.8、13.4、12.1、12.0、5.1% であった。MAPE が 10% 弱であることから、平均影響度が 10% 以上の説明変数で、かつドメイン知識の面で意外性があり [Badawy. 2019]、将来の製造条件の変更という視点で現実的な許容幅がある“造粒水温度”を、品質の変動抑制に資する潜在的因子として着目した。

Table 2. 各モデルにおける説明変数の影響度

説明変数	Model 1	Model 2	Model 3	Average
③造粒消費電力	33.0	40.1	39.9	37.7
⑤乾燥終点時の排気湿度	21.1	17.7	20.7	19.8
⑥乾燥終点時の排気温度	12.5	16.7	10.8	13.4
②造粒水温度	14.1	10.0	12.1	12.1
④乾燥時間	14.8	8.9	12.3	12.0
①スプレー速度	4.5	6.6	4.2	5.1

造粒水温度を、現在の管理温度から 6 °C 下げた場合の溶出率改善効果を、予測精度が最も高い Model 1 を用いて推定した (Fig. 2)。現行の製造条件では溶出率の最低値は 28% であったが、管理温度を下げた場合には 47% になることが推定された。また、現行の製造条件における、溶出率の最高値と最低値

の差は 44% であったが、管理温度を下げることでその差が 12% になることが推定され、品質の変動が低減することが推定された。  
今後、実際の製造設備を使い、管理温度を下げた条件で本製剤を製造し、今回推定された改善効果が得られるかどうか検証していく。

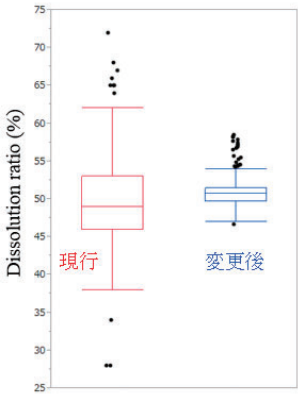


Figure 2. 造粒水温度の管理温度変更後の推定溶出率。現行の管理温度から 6 °C 下げた場合の溶出率を推定。

4. まとめ

安定的な医薬品製造を行うために、製造データを解析し製造条件を改善することは、医薬品の安定供給の観点で重要である。医薬品の製造は GMP の下で厳しく管理されているが、製造を続けていく中で品質は変動していく。この変動を抑えることができれば、より安定的な製造が可能となる。本検討では、カプセル製剤の品質パラメータの一つである溶出率に焦点をあて、機械学習の技術を適用し、モデルをスクリーニングすることで、品質変動の抑制につながる潜在的な製造因子の抽出を行なった。その結果、ドメイン知識の観点から意外性があり、製造条件の変更という視点で現実的な許容幅がある因子として、“造粒水温度”を抽出することができた。この温度を、現行の管理温度から下げることで、溶出率の変動が低減することが推定された。今後、実際の製造ラインで管理温度を変更し、今回示唆された改善効果が得られるか検証していく。併せて、できるだけ少ない実験数で、目的の検証を効率よくできる方法 (実験計画法) についても考えていきたい。

謝辞

京都大学大学院情報学研究科の加納学教授にはデータ解釈において、弊社寺邊正大研究員にはデータ解釈および資料作成においてご助言を頂きました。ここに記して謝意を表します。

参考文献

[Breiman. 2001] Leo Breiman: Random Forests. Machine Learning, Vol.45, No.1, pp.5–32, 2001.  
[Badawy. 2019] Sherif I.F. Badawy, Ajit S. Narang, Keirnan R. LaMarche, Ganeshkumar A. Subramanian, Sailesh A. Varia: Chapter 3 - Mechanistic Basis for the Effects of Process Parameters on Quality Attributes in High Shear Wet Granulation, Handbook of Pharmaceutical Wet Granulation, Academic Press, pp.89-118, 2019.