

学位論文 博士(工学)

粉体および界面活性剤配合系における
混合物特性の機械学習予測と処方設計

2026年3月

慶應義塾大学大学院理工学研究科

濱口 直

Contents

1	序論	3
1.1	本研究の目的	3
1.2	打錠機を用いた錠剤設計について	3
1.2.1	錠剤設計の概観	3
1.2.2	粉体特性と成形技術の発展	4
1.2.3	打錠装置の進歩と計測	4
1.2.4	錠剤添加剤と処方設計	5
1.2.5	錠剤の崩壊性・溶出性	5
1.2.6	品質管理と製造プロセスの高度化	5
1.3	クレンジング剤の機能と種類について	6
1.3.1	クレンジング剤の洗浄機構	6
1.3.2	油性クレンジング剤の発展	6
1.3.3	水性クレンジング剤の発展	7
1.4	本論文の構成	8
2	人工知能手法の製品開発への応用	9
2.1	本章の目的	9
2.2	マテリアルインフォマティクス	9
2.3	ケモインフォマティクス	10
2.4	プロセスインフォマティクス	10
2.5	QSAR/QSPR の研究事例	10
2.6	記述子	11
2.7	課題および本論文の目的	12
3	錠剤開発への AI の活用	13
3.1	本章の目的	13
3.2	実験手法	14
3.2.1	サンプルの作成と測定	14
3.2.2	原料粉体物性の測定	15
3.2.3	追加特徴量の設計	16
3.3	AI モデリング手法	16
3.3.1	開発データ統合とデータベース化	16
3.3.2	記述子について	17
3.3.3	データ構成	18
3.3.4	モデル構築手法	18
3.3.5	内挿予測における評価設計	19
3.3.6	外挿予測における評価設計	21
3.4	結果と考察	23

3.4.1	内挿予測におけるモデル性能	23
3.4.2	外挿予測におけるモデル性能	31
3.4.3	物理的考察	33
3.5	結論	33
4	クレンジング剤開発への AI の活用	35
4.1	本章の目的	35
4.2	実験手法	35
4.2.1	洗浄率の評価	35
4.2.2	AI モデリング手法	37
4.2.3	逆解析による処方設計と特徴量効果の可視化	39
4.3	結果と考察	40
4.3.1	AI のモデリング	40
4.3.2	逆解析による処方設計と特徴量効果の可視化	40
4.4	考察	44
4.5	結論	44
5	クレンジング剤開発への AI と分子シミュレーション手法の活用	46
5.1	本章の目的	46
5.2	実験手法	46
5.2.1	洗浄率の評価	46
5.2.2	粗視化分子シミュレーション	47
5.2.3	粗視化モデルとシミュレーションからの特徴抽出	48
5.2.4	機械学習	49
5.3	結果と考察	54
5.3.1	シミュレーションから導出された記述子	54
5.3.2	機械学習	54
5.3.3	変数の影響分析	61
5.4	結論	62
6	結論, および, 研究成果の応用と展望	65
6.1	研究目的に対する結論	65
6.2	本研究の産業応用	66
6.3	技術移転の方向性	66
6.4	今後の研究展開	67
	参考文献	69

Chapter 1

序論

1.1 本研究の目的

ものづくり企業においては、新規製品を創出して市場に届けることが企業活動の要であり、本研究はその中核工程である製品開発に焦点を当てる。製品開発では複数の特性値（スペック）を同時に満たす多目的最適化が通例であるが、その条件探索には試作・実験を伴う反復が必要となり、時間・費用の観点で大きな負担が生じている。とりわけ混合物を対象とする製品開発では、原料同士の総合作用に起因する相互作用が卓越し、自己組織化や相平衡、接着、硬化などに由来する強い非線形効果が特性を支配するため、組成の特性に対する応答は高次元・非凸・しばしば不連続となる。

本研究はこの混合物の製品開発に焦点をあて、研究対象を錠剤とクレンジング剤の二系に設定している。錠剤については、AI（機械学習）を適用し、硬度・崩壊性・溶出性・流動性など複数特性値を同時に扱うデータ駆動型の最適化課題として定式化する。一方、クレンジング剤については、一特性ではあるものの、界面活性剤混合に起因する非線形現象の核心把握を優先し、洗浄性能予測モデルの構築に取り組む。また、クレンジングにおける予測においては、CAE（Computer Aided Engineering）の技術の併用も取り組む。具体的な手法として粗視化モデルである Dissipative Particle Dynamics（DPD）を採用し、分子～メソスケールの構造情報を記述子として抽出し、AI（機械学習）と統合する。すなわち、本研究の目的は、錠剤ではAIのみで多目的設計の実効性を示し、クレンジング剤ではCAE（DPD）とAIを組み合わせることで高難度の特性予測に体系的に取り組む枠組みを提示し、配合空間の事前スクリーニングと設計意思決定の迅速化・省実験化に資することである。

1.2 打錠機を用いた錠剤設計について

1.2.1 錠剤設計の概観

経口投与製剤には錠剤やカプセルなど様々な剤形が存在するが、本章では特に打錠機によって粉末を圧縮成形する錠剤（圧縮錠）に焦点を当てる。錠剤は1回用量を正確に含有し、安定かつ持ち運びやすい形態であることから、医薬品のみならずビタミンなどのサプリメント（栄養補助食品）においても広く用いられてきた。錠剤の大量生産を可能にした打錠機の発明以来、その製造技術と錠剤設計に関する研究は、粉体工学的な基礎研究から製剤技術、装置開発に至るまで多岐にわたって発展してきた。本章では、錠剤設計に関連する粉体特性、成形技術、添加剤、打錠装置、錠剤の崩壊・溶出特性、品質管理について、技術史的・工学的観点から主にサプリメント分野を中心に概観する。

1.2.2 粉体特性と成形技術の発展

錠剤を製造するには、原料粉体として、金型への充填時に固体流動体として安定して流れ（流動性）や圧縮時に空隙を減少させ結合する（圧縮性・成形性）といった性質を備えている必要がある。多くの有効成分粉末はそのままでは打錠適性が低く、直接打錠法（粉体を造粒せずにそのまま打錠する方法）に適用できる有効成分は全体の2割未満に過ぎないと報告されている [1]。直接打錠では、粉体の物性が不十分な場合に含量均一性の不良、付着（ステッキング）や割錠（キャッピング）などの問題が生じやすく、錠剤品質に深刻な影響を及ぼす。このため、有効成分を補助賦形剤と混合した後に顆粒へと造粒し、流動性や圧縮性を向上させてから打錠する間接法が長らく主流であった。

間接法において、従来から用いられる造粒法には、結合液を加えて湿潤軟塊を形成後に乾燥・整粒する湿式造粒と、液体を用いずに圧縮応力で一次粒子を凝集させる乾式造粒（ローラーコンパクションやスラッグ法）がある [2]。例えば、乾式造粒法では、ローラーコンパクション装置により脆弱な有効成分や水・熱に不安定な物質を顆粒化して打錠適性を高めることが可能となった。一方、湿式造粒法では、高せん断混合機や流動層造粒機により結合液（例：PVPやHPMC水溶液）を添加・噴霧して湿潤軟塊を形成し、乾燥・整粒を経て流動性・圧縮性および含量均一性を向上させることができる [2]。さらに製薬分野では、気流下で微圧縮する気流式造粒、溶媒を用いずに水蒸気で湿潤させる蒸気造粒、少量の水分で誘発する湿式造粒（Moisture-Activated Dry Granulation）、発泡結合剤を用いるフォーム造粒など、新たな造粒技術が21世紀に相次いで開発されている。粉体特性を改良し成形性を確保する技術は発展を続けており、サプリメント錠剤においてもしばしば応用されている。このように、湿式造粒は複数工程を要するものの安定した造粒が可能であり、20世紀後半まで多くの製剤で採用されてきたものの、近年は有効成分や製造効率の観点から造粒を行わない直接打錠法への関心が高まり、粉体をそのまま打錠できるように物性を改善する手法も同時に発達した。

1.2.3 打錠装置の進歩と計測

錠剤の大量生産を可能にする打錠装置（打錠機）は、19世紀末に単発式からロータリープレス式へと発展し、20世紀後半には高速大量生産と精密な制御を両立する機械へと進化した。高速回転式の打錠機では、複数の金型を円盤上に配置して連続的に打錠することで生産性が飛躍的に向上し、現代では1時間に数十万錠もの製造も可能となっている。1980年代には打錠機に荷重センサや位置センサを組み込んだ計測技術が導入され、打錠圧力や充填深さをリアルタイムで監視・フィードバック制御することで、錠剤重量の自動補正や不良錠の自動排除が実現した。これにより、一錠毎の品質を監視しながらの製造（Process Analytical Technology, PAT）が可能となり、医薬品のGMP基準に適合した安定生産が可能になった。

学術研究の分野でも、打錠機の計装化は粉体圧縮挙動の解明に大きく貢献した。荷重センサで測定される上杵・下杵の圧縮力波形や、変位センサで捉える粉末の圧縮・弾性戻り挙動を解析することで、塑性変形型や脆性破砕型など材料の圧縮特性が明らかにされた。また、少量の材料で本番機の打錠サイクルを再現できるコンパクションシミュレータ装置も開発された [3]。コンパクションシミュレータは単発の試験機に高精度アクチュエータと計測系を備え、工業的なロータリープレスと同等の打錠速度プロファイルで粉末を圧縮できる装置である。この装置により、わずかな試料で圧縮挙動やスケールアップ研究が可能となり、打錠プロセスの機械的理解と開発効率向上に寄与している。近年では連続生産技術との融合も模索されており、打錠装置はさらなる高効率化・自動化が進んでいる。

1.2.4 錠剤添加剤と処方設計

錠剤には有効成分の他に様々な添加剤（賦形剤、結合剤、崩壊剤、滑沢剤など）が含まれる。添加剤は打錠に適した粉体物性を実現し、製造性と錠剤品質を確保するうえで不可欠である [4]。例えば賦形剤は錠剤の嵩と機械的強度を与える目的で用いられるが、その代表例である微結晶セルロース（MCC）は卓越した塑性変形能により高い結合力を示し、直接打錠用賦形剤として最も広く利用されていることが知られている。また、乳糖やデキストリンのような水溶性賦形剤は錠剤崩壊後の溶出液中で速やかに溶解して有効成分の放出を助ける。20 世紀末までに各種添加剤の機能解明と改良が進み、直接打錠用に単一成分の高機能グレード（改質デンプンなど）や複数成分を複合化した賦形剤が開発されてきた [5]。特に 21 世紀に入ると、高速打錠機による大量生産や直接打錠へのニーズに対応するため、複数の既存添加剤を物理的に複合化して流動性・圧縮性を大幅に改善した「共処理賦形剤」が実用化されている [6]。共処理技術により、賦形剤の多機能化（例：賦形・結合作用の両立）が可能となり、従来は造粒を要した難製錠性の成分でも直接打錠が適用できるケースが増えている。このような添加剤設計の進歩は、医薬品だけでなく有効成分含量が多く賦形剤の制約が大きいサプリメント錠剤の処方設計にも恩恵を与えている。

さらに、錠剤服用時の利便性や有効成分の溶出制御を目的として、添加剤の組み合わせによる製剤設計も発展してきた。崩壊剤は錠剤が体内で素早く崩れるよう加えられる添加剤であり、デンプンやセルロースの架橋高分子などの「超崩壊剤」は少量で高い膨潤・崩壊効果を示すため 2000 年代以降広く用いられている。これら添加剤の工夫により、近年では水なしで口腔内崩壊する OD 錠や、大量成分を含むチュアブル錠など、サプリメントにおいても様々な剤形展開が可能となっている。

1.2.5 錠剤の崩壊性・溶出性

錠剤を服用した際に有効成分が適切に放出・吸収されるためには、錠剤の崩壊性と溶出性が重要な品質特性となる。一般的な即放性錠剤では、胃腸液中でおおよそ 2.5～10 分以内に錠剤が完全に崩壊し、有効成分が溶出することが一般的な実務的目安である。崩壊は錠剤内部に浸透した水分により錠剤が粉末状に崩れる現象であり、その速度は崩壊剤の種類・量や錠剤硬度、賦形剤の水溶性などの処方要因に大きく左右される。添加剤技術の発達により、現代の錠剤では少量の超崩壊剤の添加によって硬度を保ちながら迅速な崩壊を実現することが可能となった。崩壊後の有効成分の溶出速度は、粒子径や溶解性、造粒やコーティングの有無などに依存する。医薬品分野では錠剤の溶出挙動と体内吸収（生物利用性）との相関を明らかにしようとする研究が古くから行われ、in-vitro 溶出試験による品質管理と in-vivo 実効性の関連付けが模索されてきた。しかし、従来の経験的な溶出モデルでは薬物放出機構の本質的理解が困難であり、必ずしも一般的な予測性を持たないことが指摘されている。そのため近年では、崩壊・溶出の機構論的解析や数値モデル化も進められている。サプリメントにおいては医薬品ほど厳密な溶出規格が課されない場合も多いが、最終製品の品質として適切な崩壊・溶出プロファイルを示すことが重要である点は同様であり、これら特性をコントロールする製剤設計指針は医薬品分野の知見が活用されている。

1.2.6 品質管理と製造プロセスの高度化

錠剤の品質管理は、その含有成分が規格どおり均一に含まれ、所定の機能（崩壊時間や溶出率など）と物理的性質（硬度や摩損率など）を満たしていることを保証することである。特に有効成分の含量均一性は錠剤製剤の最重要品質項目の一つであり、用量の再現性確保や安全性の観点から厳格に管理される [7]。含量均一性の不良は、混合不足による初期ブレンドの不均一や、適切に混合された粉体でも搬送・充填時の粒子分離（偏析）によって

生じ得る。このため近年の研究では、混合粉体の装置内移動やダイへの充填過程における粒径・密度の異なる粒子の分離メカニズムが詳細に検討され、偏析を抑制する装置設計や工程条件について知見が蓄積している。

また、品質管理の考え方も、従来の最終製品を検査して合否判定するアプローチ (Quality by Testing) から、製品が所定の品質になるよう製剤設計と製造条件を最適化する品質デザイン (Quality by Design, QbD) へと変化している。QbD では処方組成や工程パラメータが製品品質に与える影響を統計的に評価し、工程内での変動要因を十分に理解・管理することが求められる。このパラダイムの下、錠剤製造においても製剤因子 (粒子径分布、含水率、添加剤特性など) やプロセス因子 (混合時間、打錠圧力、打錠速度など) の製品品質への影響を系統的に調べ、設計段階で品質リスクを低減する取り組みが進められている。さらに近年では、製造中にセンサで錠剤の品質指標を連続測定し、リアルタイムにプロセスを調整する技術 (PAT) の導入も試みられている。このような先進的製造技術は、医薬品のみならずサプリメントの生産においても品質確保と効率向上に寄与するものと期待される。

以上、打錠機による錠剤設計技術の発展を、サプリメントを中心に概説した。粉体の特性改良、造粒・直打技術の進歩、高機能添加剤の開発、打錠装置の高性能化、そして品質管理手法の高度化により、現代の錠剤は安定した品質と性能を備え、大規模生産にも適した製剤形態となっている。これらの技術史的知見は今後の錠剤設計・製造プロセスの最適化にも貴重な指針を提供するものである。

1.3 クレンジング剤の機能と種類について

1.3.1 クレンジング剤の洗浄機構

クレンジング剤 (洗浄剤) の基本機構は、汚れ—基材界面のエネルギーを低下させて汚れを剥離し、流体中へ安定分散させて再付着を防ぎ、すすぎで除去することに要約される。剤型により溶解・乳化・分散の寄与は異なるが、いずれも界面活性剤など両親媒性分子の働きに基づく [8] [9] [10]。また、相似相溶 (like dissolves like) の原理により汚れは適切な相へ取り込まれて除去される。ロールアップ機構とは、疎水性表面に付着した油滴が界面活性剤等の作用で濡れ性が変化し、接触角の低下 (または変化) に伴って表面から転がり剥離する現象であり、乳化と並んで洗浄効率に寄与する [11] [12] [13]。

1.3.2 油性クレンジング剤の発展

油性クレンジング剤は、その名の通り製剤の主成分が油分で構成されるクレンジング剤であり、一般にメイクアップなどの油性汚れに対して高い洗浄力を示す [14] [15]。クレンジングオイルや油性のクレンジングクリームは、水をほとんど含まないため使用時にベタつく感触を伴うが、ファンデーションや口紅、日焼け止めなどの疎水性汚れを良く溶解し、高い汚れ除去効果を発揮する [16]。また、メイク落とし効果を客観的に評価する手法として、UV 分光に基づく評価法も提案されている [17]。実際、油性タイプのクレンジング剤は界面活性剤主体の水性タイプと比べてメイク汚れの除去効率が全般に高く、特にクレンジングオイルは最も優れたメイク落とし効果を示すことが知られている。

油性クレンジング剤の代表であるクレンジングオイルの基本的な働きは、メイクアップ製品中の油性成分を溶かし出す「溶解洗浄」である。すなわち、「油が油を溶かす」原理によってファンデーション等の疎水性汚れを溶解し浮かせる役割を担う。古典的なクレンジングクリームでは、鉱物油やワックスを主成分とした W/O (水中油滴型) エマルションを肌に塗布し、ティッシュなどで拭き取ることでメイクを除去していた。一方、近代的なクレンジングオイル製剤では、ミネラルオイルやエステル油などの溶剤性油分に界面

活性剤（乳化剤）を配合することで、使用後に水で洗い流せる処方も増加している。界面活性剤の添加により、水ですすぐ際に油分が水中で乳化され、ベタつきの少ない仕上がり感を得ることができる。典型的にはステアレス-20 や脂肪酸エステル系（PEG-7 Glyceryl Cocoate 等）といった親水基の大きなノニオン界面活性剤が用いられ、製剤中では油相に溶解しているが、水に触れると即座に O/W 型エマルジョンを形成して油污れごと水相へと運び出す [9] [18]。このような自己乳化型のクレンジングオイルは日本で 1990 年代以降急速に発展した。

さらに 2010 年代には、クレンジングオイルの洗浄力と洗い流し性を両立させるために、界面活性剤の相平衡制御に基づく高度な処方設計が報告された [19] [20]。たとえば Watanabe ら [14] は、シリコン油と水が両連続チャンネルを形成するバイコンティニュアス相を利用したメイク落とし系を報告している。一般に、界面活性剤系では、水分量の変化に伴い、逆ミセル型～両連続相～ミセル型へと相構造が遷移しうることが知られている [21] [22]。この処方では、油中に存在する界面活性剤が逆ミセルを形成して油性汚れを効果的に取り込み、水の添加によって相構造が切り替わるため、従来品と同等の高いメイク落とし効果を維持しながら、すすぎ時のべたつき残りは水性ローション並みに軽減されることが示された [21] [22]。このような高性能クレンジングオイルを実現する鍵として、界面活性剤の親水親油バランス（HLB 値）を精密に調節し低濃度領域で不要な O/W 乳化相の発生を抑えること、および極性油を添加して相図上中央付近に現れがちな液晶相を除去することが挙げられている [20]。油性クレンジング剤の設計技術はこのように発展してきており、近年では植物由来のエステル油や高分子シリコン油など新規油剤の利用や、ビタミン E 等の抗酸化剤配合によるスキンケア付加価値も志向されている。一方で、油性クレンジングは洗浄力が高い反面で肌の必須脂質まで溶出させてしまうリスクも指摘されており [10]、洗浄後の保湿ケアなどとの組み合わせを含めた検討が行われている。

1.3.3 水性クレンジング剤の発展

水性クレンジング剤は、洗浄成分として界面活性剤を主とする水系のクレンジング剤であり、一般的な洗顔料やボディソープ、クレンジングフォーム、いわゆるミセルウォーター（拭き取りタイプのクレンジング水）などが含まれる。界面活性剤は一定濃度以上で水中に会合しミセルを形成し（臨界ミセル濃度, CMC [23]）、汚れ表面への吸着による界面張力低下と、ミセルへの取り込み（乳化・分散）によって汚れの再付着を防ぐ [9] [24]。界面活性剤は分子内に親水基（極性基）と疎水基（非極性の炭化水素鎖）という 2 つの部分の併せ持つ両親媒性分子であり、この構造により水相と油相の境界に自発的に吸着して両者の混ざりにくさを緩和する [9] [24]。その結果、界面活性剤は本来固体表面に付着している油滴を表面から引きはがし（いわゆるロールアップ機構）[11]、引きはがされた油滴は多数の界面活性剤分子によって取り囲まれ安定化される（乳化・ミセル化）[10]。この油滴は水中にコロイド状に懸濁され、流水で洗い流すことで皮膚から除去される。例えば、人工皮膚上のシリコン油（Dimethicone）をモデル汚れとした実験では、アニオン界面活性剤 SLES（Sodium Laureth Sulfate）を添加すると、界面張力の低下により油滴が乳化されるが、SLES を添加しない条件では乳化は起こらない。一方、高濃度のカチオン界面活性剤 DTAB（ドデシルトリメチルアンモニウムブロミド）を添加すると、油滴表面に強く吸着して接触角が変化し、油滴が丸ごと転がり落ちるロールアップが観察される [12] [13]。このように、汚れの除去には界面活性剤の種類によって乳化とロールアップの寄与が異なり得るが、一般には界面張力を十分低下させること（乳化）と、界面活性剤の吸着による油滴剥離効果（ロールアップ）の両方が洗浄性能向上に重要である。

水性タイプの洗浄剤は高い洗浄力と泡立ちを示す一方で、皮膚への刺激や乾燥を引き起こしやすい傾向が知られている [9]。その理由の一つは、水性クレンジング剤に多用されるアニオン界面活性剤（石鹼や硫酸エステル系）の分子構造や作用特性に起因する。例

例えば石鹼は高級脂肪酸（炭素数 12～18 程度のカルボン酸）のアルカリ金属塩であり、水溶液は弱アルカリ性（pH 9～10 程度）を示す。皮膚の表面は弱酸性（pH 4.5～6）に保たれているため、高 pH の石鹼で洗浄すると皮膚表面の pH バランスが乱れ、角質層のバリア機能低下を招くことが知られている。加えて、石鹼分子は皮膚表面のタンパク質や脂質と結合しやすく、必須のセラミド等の脂質を過剰に洗い流してしまう傾向が報告されている [10]。実際、石鹼で洗浄した肌はつっぱり感や乾燥が生じやすい。一方、20 世紀中頃に登場した合成界面活性剤を用いる洗浄料（いわゆる「シンデット」）は、皮膚の天然保湿因子や脂質を極力残しつつ汚れのみを選択的に落とすことが可能で、弱酸性～中性の処方と組み合わせることで石鹼よりもマイルドな使用感を実現した。特に、アニオン界面活性剤に両性界面活性剤（ベタイン系など）やノニオン界面活性剤を併用することで刺激性を緩和した処方が一般的となっている [25]。アニオン界面活性剤は単独ではミセル径が小さく荷電密度も大きいいため角質への浸透やタンパク変性を起こしやすいが、ノニオンや両性剤との混合によりミセルの帯電を中和しサイズを大きくすることで、角質タンパク質への結合や浸透を抑制できる [26]。近年ではさらに高分子ポリマーを配合してミセル表面を覆い、ミセルが角質層に侵入しにくくする技術や、洗浄時に流出した脂質を補う疑似セラミド成分の添加など、皮膚バリアへの負荷を低減する工夫がなされている [27]。こうした改良により、水性洗浄剤においても洗浄力と低刺激性を両立した製品が増えつつある。

1.4 本論文の構成

本論文の構成は以下の通りである。第 1 章では、研究背景として錠剤およびクレンジング剤の機能と種類を概説し、本研究の位置づけを述べる。第 2 章では、製品開発における人工知能手法（AI）の応用として、マテリアルインフォマティクス、ケモインフォマティクス、プロセスインフォマティクス、QSAR/QSPR および記述子の考え方と課題・展望を整理する。第 3 章では、錠剤開発への AI の活用として、多目的特性予測に焦点を当て、データセット構築、記述子設計、機械学習・深層学習モデルの設計、評価方法、さらに外挿予測について述べる。第 4 章では、クレンジング開発への AI の活用として、洗浄率予測モデルの構築、分子記述子の活用、および、性能評価と実測検証を示す。第 5 章では、AI と分子シミュレーション（DPD）を統合したクレンジング配合予測の手法を提示し、シミュレーション由来特徴量、モデル性能、変数影響度を考察する。最後に第 6 章で、結論ならびに研究成果の産業応用・技術移転の方向性と今後の研究展開を述べる。

Chapter 2

人工知能手法の製品開発への応用

2.1 本章の目的

近年、人工知能（AI）は社会の様々な分野で革新的な成果をあげており、製品開発プロセスにも大きな影響を及ぼしている。中でも、機械学習や深層学習に代表されるデータ駆動型の手法は、大規模データから人間には見出しにくいパターンや関係性を学習し、設計や意思決定を高度化する手段として注目されている。本章では、材料・化学分野を中心に、機械学習手法の製品開発への応用に関する最近の動向を概説する。とりわけ、材料科学におけるマテリアルインフォマティクス（Materials Informatics; MI）、化学・創薬分野のケモインフォマティクス（Cheminformatics）、および製造プロセス最適化のためのプロセスインフォマティクス（Process Informatics）に焦点を当て、さらにそれらの基盤技術として重要な QSAR（Quantitative Structure-Activity Relationship: 構造－活性相関）や QSPR（Quantitative Structure-Property Relationship: 構造－特性相関）手法と記述子（descriptors）の概念について論じる。以下では、それぞれの項目を順に説明し、製品開発への具体的な応用事例や課題・展望を述べる。

2.2 マテリアルインフォマティクス

マテリアルインフォマティクスとは、材料に関する実験データや知見をデータ科学的手法で活用し、新材料の探索や材料特性の予測を行う研究分野である。例えば「材料に関する科学・工学データを計算手法によって処理・解釈することで材料研究に役立てる」と定義されており、生物学や創薬分野で定着した手法を材料分野に展開する形で 2000 年代から発展してきた [28]。2011 年に米国で Materials Genome Initiative が提唱された後、2015 年頃からマテリアルインフォマティクス研究の出版数は急増しており、大規模データの蓄積と深層学習技術の成功がこの流れを後押しした [28]。機械学習は現在、材料特性予測の主要な手法となっており、密度汎関数理論（DFT）計算など従来の物理シミュレーションに比べて圧倒的に短時間で物性を推定できる [28]。実際、MI 手法により合金や高分子、セラミックスなど様々な材料系で新規材料候補が探索・設計されていると報告されている。また、深層学習を用いた QSPR モデルにより、有機太陽電池材料のエネルギー変換効率を高精度に予測することにも成功している [29]。Wu らによるこの研究では、多層パーセプトロンからなるディープニューラルネットワークを物性予測モデルに適用し、エネルギー材料の性能を効率良く評価できることを示した。さらに、物理シミュレーションとデータ駆動型モデルを組み合わせたアプローチも取られている。例えば、分子動力学シミュレーションから得られる挙動指標（イオンクラスタリング等）を記述子として、電解質の輸送特性を機械学習で予測する研究が報告されている [30]。このように、マテリアルインフォマティクスはデータと機械学習を駆使して材料開発を加速する強力な手段となっている。

2.3 ケモインフォマティクス

ケモインフォマティクスは、化学分野における情報技術活用の総称であり、特に創薬を中心とした分野で発展してきた。1998年に初めて提唱された定義では「創薬プロセスにおいて、データを情報に、情報を知識に変換し、より迅速かつ的確な意思決定を行うための手法」と説明されている [31]。このようにケモインフォマティクスは医薬品探索に端を発するが、その後対象は薬理活性に留まらず、化学構造のデータベース管理、化合物の仮想スクリーニング、化学空間の多様性解析など、幅広い方法論を含む学際領域へと発展した [31]。実際、化合物の大量合成とハイスループットスクリーニングが普及した 1990 年代以降、蓄積される構造-活性データから有用な知見を得るために機械学習によるデータマイニングや仮想評価が不可欠となっており、それがケモインフォマティクスの中心的手法の一つが、QSAR である。QSAR は分子の構造情報（化学構造や物性パラメータ）からその生物活性や毒性を予測するアプローチで、創薬初期のヒット化合物の評価・最適化や環境中の化学物質リスク評価に広く用いられてきた [31]。一方、物性値の予測に焦点を当てた QSPR も材料開発や化学プロセス設計で重要な役割を果たしている。

2.4 プロセスインフォマティクス

プロセスインフォマティクスは、材料開発と製造プロセスをデータ駆動型で統合的に最適化しようとする新しい概念である。すなわち、「何を作るか」（目標とする材料特性）と「どう作るか」（合成・成形などの工程条件）を情報学的手法で密接に関連付け、製造プロセス全体を包括的に最適化しようとする試みである [32]。マテリアルインフォマティクスの発想を製造プロセス（プロセス変数）まで拡張したものであり、機械学習によるプロセス制御や最適化技術とも位置付けられる。Funatsu によって 2010 年代後半に提唱されたこの概念は、製造業で従来経験則や個別最適で行われてきたプロセス設計を変革しようものとして注目されている [32]。具体的には、材料の合成・加工条件（温度、圧力、濃度など）やリアルタイム計測データ（例えば分光法による反応モニタリング）を記述子としてモデルに組み込み、製造中の材料構造や特性変化を予測・制御することを目指す [32]。これにより、目的の特性を持つ材料を安定に製造できるプロセス条件をデータに基づき導出することが可能となる。実際、プロセスインフォマティクスの考え方により、品質予測のためのソフトセンサー開発やマルチモーダルデータ統合モデルの構築が進められている [32]。このような包括的アプローチは、新材料開発に新たな視点と効率化をもたらすと期待されている。

2.5 QSAR/QSPR の研究事例

QSAR (Quantitative Structure-Activity Relationship) および QSPR (Quantitative Structure-Property Relationship) は、分子の構造的特徴とそれに由来する応答（生物活性または物性）との間の関係を数理モデルとして表現する手法である [31]。1960 年代の Hansch の先駆的研究以来、QSAR は医薬化学の分野で薬効や毒性の予測に活用されてきたが、現在では化学・材料分野全般で、未知物質の特性を推定するための基本的手段となっている。初期の QSAR/QSPR では、疎水性や電子的性質など少数の分子物性パラメータと目的変数（活性・物性値）との相関を多重線形回帰分析する手法が中心であった。しかし近年では、より多様な記述子を扱うために部分的最小二乗法 (Partial Least Squares; PLS) や機械学習による非線形モデルが導入されている。例えば、Absalan ら (2004) は遺伝的アルゴリズムによる変数選択と PLS 回帰を組み合わせ、双頭界面活性剤（ジェミニ型界面活性剤）のミセル会合に関する構造-特性関係を解析している [33]。Mallakpour ら (2010) は、

分子構造から理論計算で得られる複数の分子記述子を入力とし、ポリマーの固有粘度を予測する QSPR モデルを人工ニューラルネットワーク (ANN) により構築した [34]. また, Khan ら (2018) は 2 次元記述子を用いた回帰モデルにより, 多様な高分子化合物の屈折率を高い精度で予測することに成功している [35]. 安全性・環境分野への応用例として, 有機化合物混合物の引火点を QSPR で予測する研究 [36] や, 多成分系の蒸気液相平衡に関わる臨界体積を推算するモデル [37] も報告されている. 環境の観点では, 企業由来の新興汚染物質による環境リスクが指摘されており [38], 水生生物への影響評価のための指標も整備されつつある [39]. これに対し, 複数化学物質の混合毒性を予測する QSAR モデルも開発されている. Chatterjee ら (2021) は, 二成分混合化学物質の水生毒性について, 各成分の 2 次元構造記述子から QSAR モデルを構築し, PLS 回帰により混合による影響を良好に予測できることを示した [40]. 医薬品の創薬過程でも QSAR は欠かせない技術であり, 多数の候補化合物の活性や毒性を仮想評価することで実験的を絞り込むことができる [31]. さらに, 医薬品製剤設計の領域にもデータ駆動型予測が応用されている. Hayashi ら (2019) は, 薬剤錠剤の密度を構成成分の物性情報から予測する機械学習モデル (QSPR) を報告しており, ブーステッドツリーモデルにより実測値と高い一致を示している [41].

モデル構築に用いるアルゴリズムも高度化している. 線形回帰やその派生の PLS だけでなく, 決定木 (例えば勾配ブースティングツリー), サポートベクターマシン, ニューラルネットワークなど多様な機械学習手法が QSAR/QSPR に導入されている. 例えば, 前述の Mallakpour らの研究では ANN が活用されており [34], Abbasi ら (2014) は多層パーセプトロンおよび適応神経ファジィ推論システム (ANFIS) を用いて炭化水素二成分混合液の拡散係数を予測している [42]. また, Hayashi らは製剤密度予測に勾配ブースティングツリーを適用してモデルの精度向上を達成している [41]. 深層学習の応用例も徐々に増えてきており, Wu ら (2020) はディープラーニングを用いた QSPR モデルによって有機薄膜太陽電池のエネルギー変換効率を高精度に予測できることを示した [29]. このような最新手法まで含め, 混合物系への QSAR/QSPR 適用には独特の課題がある. Muratov ら (2012) は混合化学物質の QSAR 手法について総説し, 成分間の相互作用 (相乗効果や拮抗効果) を考慮する必要性を指摘している [43]. さらに最近の動向として, 多成分系の QSAR/QSPR 研究を概観し, データセット整備や記述子・アルゴリズム開発の方向性を整理したレビューも報告されている [44]. 例えば, グラフニューラルネットワーク (GNN) により界面活性剤混合溶液の温度依存 CMC を予測する枠組みが提案されている [45]. また, 処方設計 (配合設計) に対しても, 機械学習を用いて製品設計を支援する試みが報告されている [46]. このように, QSAR/QSPR 手法は化学物質の機能・特性予測に広く応用されており, 化粧品から医薬品, 材料まで様々な製品開発に貢献している.

2.6 記述子

記述子とは, 分子の構造や性質を数値化して表現した指標であり, QSAR/QSPR モデルの入力変数 (特徴量) として用いられる. 分子量や極性表面積のような簡単なものから, トポロジー指数や 3 次元形状記述子, さらに量子化学計算から得られる電子分布指標まで, 多種多様な記述子が提案されている. モデルの予測性能は用いる記述子の情報内容に大きく左右されるため, 目的に応じて適切な記述子を選定・生成することが重要である. 例えば, Mallakpour らの高分子固有粘度 QSPR 研究では, 分子構造から理論計算で導出された複数の分子記述子がモデル構築に用いられている [34]. 複雑な系, 特に混合物系に対しては, 各成分の記述子から混合全体を特徴付ける新たな「混合記述子」を設計する必要がある. Faramarzi ら (2019) は, 二成分混合物の沸点を予測する QSPR モデルにおいて, 構成成分それぞれの物性値や構造情報を組み合わせた独自の混合記述子を提案している [47]. また, Wang ら (2018) は「仮想的な混合物質」の記述子を導入し, 各成分の記述

子から混合溶液の毒性を予測する手法を報告している [48]. このような工夫により, 混合物における成分間の非線形な影響も定量化できる.

記述子は数百~数千種類にも及ぶ場合があり, モデル構築では不要な記述子を削減するための特徴選択や次元圧縮も用いられる. また, モデルの解釈性を高めるため, どの記述子が予測結果に寄与しているかを分析することも重要である. Lundberg ら (2017) による SHAP (SHapley Additive exPlanations) の提案は, 複雑な機械学習モデルにおいて各特徴量の寄与度を定量評価する汎用的手法であり, QSAR/QSPR モデルの解釈にも応用されている [49]. 近年では, 深層学習の発展により人手で記述子を設計せずとも, モデル自体に特徴表現 (潜在変数) を学習させるアプローチが現れている. グラフニューラルネットワーク (GNN) は, 化合物をノード (原子) とエッジ (結合) からなるグラフ構造として入力し, ニューラルネットワークによって分子構造の特徴ベクトルを自動学習する手法である. Wang ら (2023) は, 多成分混合系の物性を予測する高効率な GNN モデル MolSets を開発し, ポリマー/界面活性剤ブレンドなど複雑な混合材料の性質を精度良く予測できることを示した [50]. さらに, 自己教師あり学習を用いた分子グラフ表現の事前学習が進展しており, 下流タスクの汎化性能向上に寄与することが示されている [51]. このように, 記述子設計の自動化や高次元特徴量の活用が進むことで, QSAR/QSPR モデルの適用範囲と精度は着実に向上している.

2.7 課題および本論文の目的

これまで述べた既存手法については課題も存在する. 例えば, 混合系として二成分混合・単一特性を対象とするものも多く, 実開発で支配的な「多成分・多特性」系とはスケールが異なる場合がある. また, 大規模データ ($n \geq 10^3$) を前提とする枠組みも多い一方, 少量の実製剤データに対する体系的検証は限られる. さらに, 内挿領域での予測事例が中心で, 製品開発で不可避な外挿領域における精度評価や信頼度提示の整理は十分とはいえない. 加えて, 粉体物性・粒度分布・メソスケール自己組織化などの実務的・物理的指標を統合した例もまだ限られる. それらの課題を踏まえ, 本論文の研究目的を以下に設定する.

1. 少量データかつ外挿条件に対応する予測モデルと精度評価の枠組みを整備し, 不確実性 (信頼度) を定量化して提示する (第 3 章, 第 5 章).
2. 物性整合的な記述子体系として, 粉体物性・粒度分布 (第 3 章), 分子記述子 (第 4 章), および DPD/粗視化モデリング由来の指標 (第 5 章) を橋渡しして統合し, 予測能力と説明性の両立を図る.
3. 実務スクリーニングに利用可能なモデル枠組みを構築し, 候補配合の事前評価と意思決定を支援することで, 試作の優先順位付けと知見の形式知化を促進する (第 3 章~第 5 章).

以上を段階的に統合することで, 開発工程の試作回数の削減と知見獲得を図り, データ駆動による自律的で持続可能な混合系の処方設計を実現を目指す.

Chapter 3

錠剤開発への AI の活用

3.1 本章の目的

本研究の目的は、錠剤（サプリメント）開発における複雑な最適化課題を解決するため、AI 技術を活用した統合的な開発支援システムを構築することである。錠剤開発では、複数の原料を組み合わせ、各成分の配合比やスペック（硬さ、崩壊性、溶けやすさ、飲みやすさ、コスト等）を同時に最適化する必要がある。この最適化プロセスにおいて、原料の組み合わせや配合比の選択肢が膨大であるため、全てを実験で網羅することは現実的ではない。さらに、開発担当者ごとにノウハウや実験データが分散し、過去の失敗や他者の知見が十分に活用されていないという組織的な課題も存在する。

これらの課題を解決するため、本研究では過去の実験データや原料特性を統合・学習することで、最適な配合やスペックを提案できる AI システムの構築を目指した。具体的には、機械学習技術を活用して、原料の物理化学的特性、配合条件、および最終製品の品質指標の間の複雑な関係性をモデル化した。このアプローチにより、開発期間の短縮、コスト削減、製品品質の向上を目指した。

加えて、本研究の特徴は、実際の開発現場で活用可能な実用的な AI システムを構築することにある。単なる理論的な最適化手法の提案ではなく、既存製品の課題解決や新規価値創出に直接的に貢献するシステムの開発を目指す。これにより、製薬・サプリメント業界における開発効率の向上とイノベーションの促進に寄与することを最終的な目標とする。

上記を踏まえ、本章の目的を以下に設定する。

- 各特性値を予測する AI モデルにより、どの程度の予測精度が得られるかを定量的に明らかにする。
- 後述する粉体物性情報および粒度分布情報の追加が、予測精度の向上に寄与するかを検証する。また、精度の再現性・頑健性を検証するため、乱数シードによるデータ分割の揺らぎに対する性能の安定性もあわせて確認する。
- 錠剤設計においては、新規原料利用・原料代替・スペック改善などの目的故に、積極的に外挿領域を探索することも多い。この開発課題を反映し、外挿領域での予測に対して記述子情報がどの程度有用かを明らかにする。
- 外挿予測に関しては、適用範囲（Applicability Domain; AD）に基づき、層別での性能差とカバレッジを定量化し、外挿リスクを評価する。
- 得られた結果に対する物理的妥当性の確認を通じて、予測モデルを検証する。

3.2 実験手法

3.2.1 サンプルの作成と測定

錠剤サンプルは、賦形剤、結合剤、崩壊剤、滑沢剤、その他の添加剤で構成される。使用した原料の例を表 3.1 に示す。なお、造粒末については、原料粉体と同様に原料として扱った。また、サンプル作成工程として、各原料を精密天秤で秤量し、配合処方に従って均一に混合する。その後、混合粉末を試験用単発式打錠機（OKADA SEIKO 社製 N-30E）にて充填・打錠し、設定の圧縮荷重で圧縮成形して錠剤を製造した。なお、圧縮荷重は 4.4～19.6 kN（450～2000 kgf）とした。

表 3.1: 本研究で調製した処方に使用した原料例

Category	Number of materials used	Examples
Fillers (Diluents)	8	CEOLUS UF-711 CEOLUS UF-702
Disintegrants	9	PCS-FC30 SUNROSE F-20LC
Polyols (Sugar alcohols)	4	AMALTY MR-50 Powdered Maltitol G-3
Lubricants	4	Calcium stearate Magnesium stearate
Binders	2	METOLOSE NE-4VF CELNY SSL SFP
Glidants	2	SYLOPAGE 720

本研究では、錠剤開発に関わる評価指標として、打錠後の錠剤物性（硬度・崩壊時間・厚み）と、打錠前粉体の流動・凝集特性を含む 5 つの目的変数を予測対象とした。以下に各特性値の説明と測定方法を記述する。これらの目的変数は、錠剤の性能・品質・製造適正（充填・打錠安定性を含む）を評価する上で重要な指標であり、錠剤設計においても重要な指標となる。なお、本章での硬度・厚み等は混合処方として作製した錠剤の測定値である。また、各目的変数によってデータ取得数も異なるため併記する。

- 硬度（Hardness, $n = 1209$ ）：錠剤が外力に耐える機械的強度を表す特性値であり、輸送・取り扱い時のチップングやキャッピング、製造ラインでの欠け・割れの発生に直結する。一方で過度な硬度は崩壊性・溶出性を阻害しうる。デジタル硬度計（FUJIWARA 製 KHT-40N）にて測定。
- 崩壊時間（Disintegration Time, $n = 882$ ）：錠剤が規定条件下で崩壊するまでの時間を示す特性値であり、特に即放性製剤では初期溶出性・吸収性に影響する。崩壊剤の種類・量、潤滑剤量、硬度・孔隙率、粒子径分布や親水性賦形剤の影響を強く受ける。本研究においては、崩壊試験器（TOYAMA 産業社製 NT-600）にて、37℃水中で測定。
- 流動性指標（Flow Function, $n = 151$ ）：粉体がせん断下でどの程度自重により流動するかを表す特性で、白への均一充填や重量ばらつき（含量均一性の前提）に直結する。粒度分布・粒子形状・表面粗さ・含水率・帯電、ならびに滑沢剤・流動化剤の添加によって変動する。設計段階では造粒方式・結合剤選択や粒度設計で目標流動性を確保し、ホッパ角度・フィード速度など工程設定に反映する。ブルックフィールド社製パウダーフローテスターにて測定。

- 凝集性 (Cohesion, $n = 145$) : 粒子間に働く付着・凝集の強さを示し、高い場合はブリッジングやラットホール、分級・偏析の誘因となり流動性を低下させる。粒子径の微細化、含水率の偏り、表面エネルギーや静電気の影響を受けやすい。配合設計では粒度の最適化、グラニュレーション、グライド剤付与等で凝集を制御し、安定した打錠挙動と充填均一性を確保する。ブルックフィールド社製パウダーフローテスターにて測定。
- 厚み (Thickness, $n = 60$) : 打錠後の厚みは充填量と圧縮度合いの結果として現れる指標で、飲みやすさ、崩壊性・溶出性、製造安定性に影響する。重量・硬度・孔隙率（ひいては崩壊・溶出）と相関を持ち、表面積/体積比の変化により薬物の拡散挙動も変動する。金型・杵の仕様や圧縮条件に依存し、規格内管理は外観適合性・包装適合性・製造安定性の観点でも重要である。マイクロメータ (Mitutoyo 社製 PK-1012CPX) にて測定。

3.2.2 原料粉体物性の測定

後述の記述子に活用するため、原料の物理的・化学的性質を以下のように測定した。

- ゆるめ嵩密度 (Loose Bulk Density) : 粉体の充填性を表す指標であり、粉体の流動性評価に用いる。100 mL のメスシリンダーに粉体を投入し、すり切り後に粉体質量 m を精密天秤で測定し、 $\rho_{\text{bulk}} = m/100 \text{ mL}$ として算出。
- タップ嵩密度 (Tap Density) : タッピング後の充填状態における嵩密度。100 mL メスシリンダーに粉体を投入し手動タッピング後、充填体積 V_{tap} を読み取り、 $\rho_{\text{tap}} = m/V_{\text{tap}}$ として算出。
- 圧縮度 (Compressibility; Carr's index) : ルーズ状態とタップ状態の差に基づく流動性指標。

$$\text{CI} = 100 \times \frac{\rho_{\text{tap}} - \rho_{\text{bulk}}}{\rho_{\text{tap}}}$$

値が大きいほど圧縮性が高く、流動性は低下する傾向を示す。

- Hausner 比 (Hausner Ratio) : ルーズ状態に対するタップ状態の密度比。

$$\text{HR} = \frac{\rho_{\text{tap}}}{\rho_{\text{bulk}}}$$

値が大きいほど粒子間相互作用が強く、流動性は低い傾向を示す。

- FLODEX: Hanson Research Corporation 社製 FLODEX を用い、所定オリフィスからの自重排出性（開始可否・連続性）を評価し、粉体の流動挙動の一指標として用いた。
- 乾燥減量: Mettler Toledo 社製 HX204 を用い、80 °C・20 分の条件で加熱処理前後の質量差から算出。
- 溶解性スコア (Solubility score) : 水に対する溶解挙動を攪拌下で目視観察し、溶解／難溶／不溶を判定した指標。
- 脂溶フラグ (Lipophilic score) : 原料の溶解挙動に基づき、脂溶性/非脂溶性を判定した指標。

- ・流動性スコア（Flowability score）：原料粉体の流動性指標の測定結果。
- ・粒度分布：レーザー回折散乱法粒度分布計（HORIBA 製 LA-960）を用い、分散エア圧 0.4 MPa 条件で測定。分布データとして、 $0.115\mu\text{m}$ から $5000\mu\text{m}$ までの 80 点の共通粒子径グリッドに対応する頻度分布を抽出した。

3.2.3 追加特徴量の設計

本研究では、上記で算出した原料単体スカラー特性（粉体物性・スコア）をベースとしつつ、混合処方側の幾何学・打錠条件・組成バランスを反映するために、追加特徴量を設計した。

- ・R 部高さ：錠剤の曲率半径に基づく凸形状の寄与を表す形状パラメータであり、平面形状のみでは表せない体積・密度の差を反映することを意図した。
- ・R 部体積：上記 R 部高さと同錠剤径から得られる凸部の体積寄与の代理指標であり、形状差が体積（ひいては密度・空隙）に与える影響を反映することを意図した。
- ・空隙度：重量と推定体積から得る「詰まり具合」の目安であり、硬度や崩壊に関わる密度・空隙構造の違いを特徴量として与えることを意図した。
- ・1/重量：重量に対する非線形性（例えば小型錠と大型錠でのスケール差）を拾うための逆数特徴とした。
- ・打錠圧（単位面積換算）：錠剤面積で正規化した面圧の指標であり、直径の異なる錠剤間でも圧縮応力の強さを比較可能にすることを意図した。
- ・原料カテゴリ比率（群和）：混合状態を反映するために、錠剤重量で正規化した 9 つの比率（結晶セルロース割合、崩壊剤割合、糖アルコール割合、滑沢剤割合、結合剤割合、流動化剤割合、顆粒割合、主成分割合、その他割合）を算出した。
- ・2 次交互作用（interaction-only）：上記の 9 つの原料カテゴリ比率のうち、結晶セルロース割合から顆粒割合までの 7 比率について、interaction-only（二乗なし）の 2 次交互作用項を作成し、合計 21 次元の交互作用特徴を追加した。

3.3 AI モデリング手法

図 3.1 にモデリングのフロー図を示す。

3.3.1 開発データ統合とデータベース化

本研究の事前準備として、各研究員が独自の Excel で分散管理していた処方・プロセス・測定データを共通フォーマットへ統一し、RPA（Robotic Process Automation）を用いた自動取り込み（検証→変換→格納）の仕組みを構築して内部データベース (DB) として再編した。あわせて利用原料の物性情報を格納した物性 DB も作成し、両 DB 間は原料 ID 体系で参照一貫性を確保した。運用上は、共通テンプレートから RPA が必須項目の充足・単位／表記ゆれの正規化・ID 整合・重複検出を行い、ETL パイプラインでスキーマに準拠して格納する。また、スキーマ定義に加え、単位の標準化、欠損・重複の取扱いルール、由来メタデータ（製品名・時期・装置・測定者・版）付与、更新ログによる版管理を整備した。

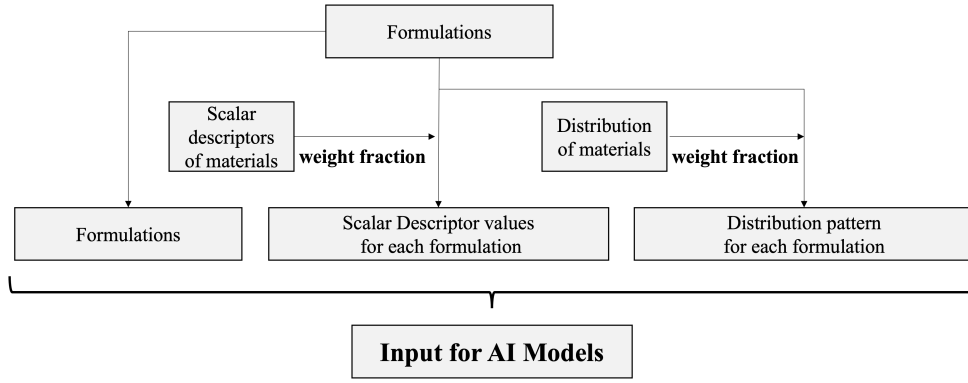


図 3.1: AI モデリングのフロー図

本研究では以降、当該 DB 群をモデル学習の単一ソースとして用いている。なお、DB 化に伴い副次的に、研究者横断の検索・共有・進捗可視化が容易となり、重複試験の抑制やリーダーマネジメントの効率化、ならびに再現性・トレーサビリティの向上が得られた。

3.3.2 記述子について

3.2.2 で測定した粉体物性値を、錠剤の物性値予測に必要な記述子として利用した。ただ、各記述子は各原料の値である一方、予測したい対象は混合物である。そこで、以下の式に基づき、各原料の混合記述子を算出した。

$$\mathbf{p}_{\text{mix}} = \sum_{i=1}^n x_i \cdot \mathbf{p}_i \quad (3.1)$$

ここで、 $\mathbf{p}_{\text{mix}}$ は混合物の記述子ベクトル、 x_i は成分 i の質量分率、 $\mathbf{p}_i$ は成分 i の記述子ベクトルであり、 $\sum_{i=1}^n x_i = 1$ の正規化条件が成り立つ。この定式化は、混合物レベルの特徴量（混合記述子）を構成するための実務的な近似であることを期待している。例えば、本研究で扱う粒度分布のようなベクトル情報も、同様に配合比に基づき処方分布（混合粒度分布）を構成できる。

このとき、記述子がスカラーの場合、 $\mathbf{p}_i$ はスカラー値である。本研究では粒度分布が元データとしてはベクトル（80 次元）であるが、データ数に対して高次元となるため、後述する通り、まず配合比に基づき混合粒度分布を構成し、その後に要約特徴へ変換して次元削減した。

一方、粒度分布以外の粉体物性値は全てスカラーである。なお、スカラー記述子の中には、分析できない、もしくは、過去保有していた原料のため再分析ができないなどの理由により欠損値があった。そのため、これらの欠損値に対しては、 k -近傍法補完器 ($k = 5$) により欠損値補完を行った。また、混合後の粉体を想定した追加特徴量（原料カテゴリ比率等）の定義は 3.2.3 に示したとおりである。

本研究においては、目的変数ごとに採用する記述子を変更した。例えば、ゆるめ嵩密度・タツプ嵩密度・圧縮度（Carr's index）・Hausner 比・乾燥減量などの基本粉体物性は全目的変数で共通に含めた。硬度および崩壊時間、厚みでは、形状・密度・応力の影響を反映する追加特徴量（R 部高さ/体積、空隙度、1/重量、打錠圧）を追加した。一方、流動性指標および凝集性では FLODEX および流動性スコアを、崩壊時間では溶解性スコアおよび脂溶フラグを追加した。これらの操作の結果、追加されるスカラー記述子ブロックの次元数は目的変数により異なり、本研究では 37–47 次元（硬度:46, 崩壊時間:47, 流動性指標:37, 凝集性:37, 厚み:45）であった。

粒度分布は原料の分布ベクトル（80 次元）を配合比で式 3.1 に則って加重平均して処方ごとの混合粒度分布を構成した．この時，粒度分布情報が欠損する原料については 0 ベクトルとして扱うことで処方分布に寄与しない処置をした上で，欠損しない原料の合計が割合が 100% となるように正規化した．その後，過学習リスクを抑えるために要約特徴（80 次元から 10 次元）へ変換して次元削減した．具体的には，分位点 (d_{10}, d_{50}, d_{90})，平均粒径，標準偏差，歪度，尖度，粒度 $25\mu\text{m}$ 以下の粉体割合，粒度 $150\mu\text{m}$ 以上の粉体割合，およびエントロピーを特徴量として用いた．

3.3.3 データ構成

前述の記述子が精度向上に効果があるかを検証するため，以下 3 種類のデータセットを構築した．まず，原料の添加率とプロセス情報（打錠圧・杵種類・錠剤径など）を連結した 157 次元の特徴量（MP：Materials + Processes），次に，MP に 3.3.2 で述べたスカラー記述子ブロック（目的変数ごとに 37–47 次元）を追加した特徴量（MPD：Materials + Processes + Descriptors），最後に，MPD に対して，粒度分布由来の要約特徴（約 10 次元）を追加した特徴量（MPDD：Materials + Processes + Descriptors + Distribution）を設定した．粒度分布の混合・正規化，および要約特徴への変換手順は 3.3.2 に示した通りである．

3.3.4 モデル構築手法

3.3.4.1 モデル選択

原料添加率情報，記述子，および，粒度分布情報を含む全特徴量の次元数は目的変数によって異なるが，高次元（MPDD ではおおむね 200 次元台）となるため，これらの多数の特徴量から物性値を予測する法則性を獲得するために機械学習と深層学習を適用した．また，本研究の目的変数はすべて連続値であるため，教師あり学習（回帰）を選択した．このとき，入力となる説明変数は 3.3.3 で説明したデータであり，目的変数は 3.2.1 で説明した錠剤開発に関わる評価指標（打錠後の錠剤物性：硬度・崩壊時間・厚み，および打錠前粉体特性：凝集性・流動性指標）である．

本研究の機械学習において，2 つの決定木ベースモデル（Random Forest Regressor (RF) と Extremely Randomized Trees (Extra-Trees, ET) Regressor），2 つの線形ベースモデル（Lasso と Partial Least Squares, PLS），1 つのサポートベクターベースモデル（Support Vector Regressor, SVR）を採用した．また，深層学習モデル（Neural Network, NN）も用いたが，3.3.3 で説明したデータパターンを別々のブランチへの入力として投入する設計とした（図 3.2 参照）．具体的には，原料情報とプロセス情報（157 次元），スカラー記述子情報（目的変数ごとに 37–47 次元），粒度分布の要約特徴（約 10 次元）の 3 つの入力ブランチを設けた．

この構造において，粒度分布については，まず混合粒度分布から要約特徴へ変換して次元削減した上で，それらの相互作用を加味するため，この入力層用の深層学習レイヤーにより処理を行った．同様に，スカラー記述子セットについても，この入力層用のレイヤーにより処理を行うことで，スカラー特徴量同士の相互作用特徴量抽出を行う．各入力ブランチは独立した深層学習層を通して処理され，その後，これらのデータと原料とプロセスデータ（図 3.2 における”Materials and Processes”）を連結し，再度深層学習層を通して，目的変数を予測した．

機械学習，深層学習ともにモデルを目的変数の数だけ独立に構築し，各物性値に対して最適化された予測モデルを構築した．また，説明変数のうち配合比（原料添加率）は標準化せず，それ以外（プロセス条件，スカラー記述子，粒度分布由来の要約特徴など）の数値特徴量を平均 0，標準偏差 1 に標準化して適用した．一方，粒度分布は各サンプル

(行)ごとに合計が1となるように正規化した混合粒度分布から要約特徴を算出しており、これらはスカラー特徴として他の連続特徴量と同様に扱った。

なお、後述する内挿範囲の予測に関しては上記のすべての目的変数・モデルを用いたが、外挿範囲予測においては、データ数の観点から予測対象目的変数は硬度 (Hardness) と崩壊時間 (Disintegration Time) に絞った。また、外挿評価 (3.3.6.1 で述べる 5-fold 相当の Rolling-origin) における外挿比較では、NN は学習のばらつきが大きく、少数 fold の設定下で評価値の外れ値を生じやすいことが推定されたため、本章の外挿比較からは除外した。

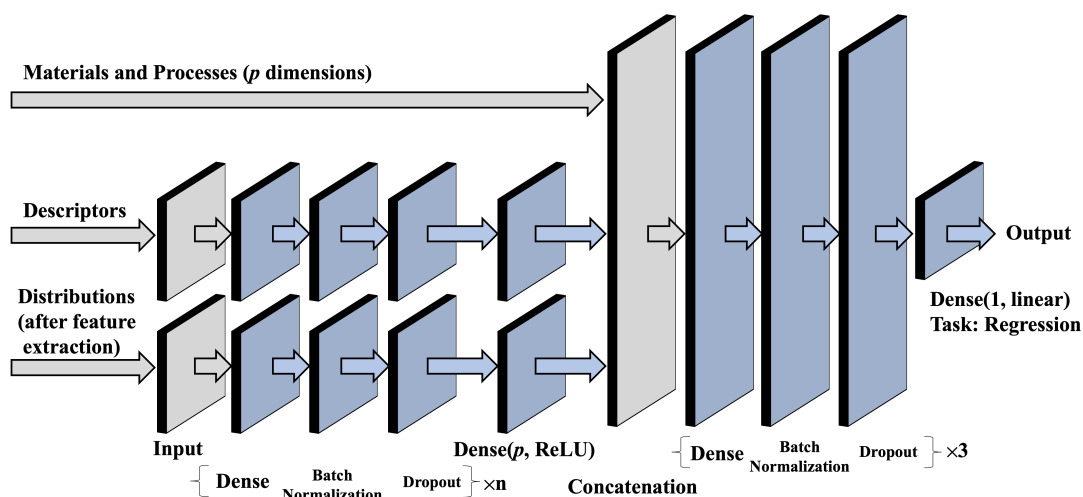


図 3.2: 深層学習の構成図

3.3.4.2 ハイパーパラメータの最適化

機械学習モデルにおけるハイパーパラメータ最適化では、各モデルのハイパーパラメータ範囲を表 3.2 に示す。ハイパーパラメータはグリッドサーチ法に基づいて最適化した。一方、深層学習モデルのハイパーパラメータは、計算時間増大が発生するため、ベイズ最適化手法 (Optuna) により検証損失 (RMSE) の最小化を目的として探索した。深層学習における探索対象ハイパーパラメータの種類およびその探索範囲を表 3.3 に示す。各試行では、同一の前処理・特徴量構成に基づくモデルを生成し、検証損失に基づく早期終了および学習率減衰を併用して学習し、検証 RMSE を目的関数値として評価した。Optuna による試行回数は損失値の収束を確認しつつ 50 回とした。

3.3.5 内挿予測における評価設計

3.3.5.1 データ分割

内挿予測は、観測済みの設計空間内での汎化性能と記述子利用の純粋な効果を共変量シフトの影響なく評価するためのものとして設定した。機械学習および深層学習モデルにおける精度評価では、データの適切な分割と再現性の確保を重視した評価手法を採用した。内挿評価においては、全データを Train/Development(Dev)/Test に 50%:30%:20% の比率で分割した。Train データで学習、Dev データにて検証とハイパーパラメータ選定を行い、その後 Train+Dev データでの再学習を行った後、Test データでの最終的な汎化性能を評価した。また、分割乱数による依存性を低減するため、分割乱数シードを変更しつつ、反復実験 ($n = 50$) を実施し、全試行の平均値を評価指標として採用した。

表 3.2: 機械学習におけるハイパーパラメータの探索範囲

Model Name	Hyperparameter	Value
Extremely Randomized Tree Regressor	Number of Estimators	50, 100
	Maximum Features	0.3, 0.5, 0.7
	Maximum Depth	None, 10, 20
	Minimum Samples per Leaf	1, 5, 10
	Minimum Samples for Split	2, 10
RandomForestRegressor	Number of Estimators	50, 100
	Maximum Features	0.3, 0.5, 0.7
	Maximum Depth	None, 10, 20
	Minimum Samples per Leaf	1, 2, 4
	Minimum Samples for Split	2, 5
SVR	Regularization Parameter (C)	0.001 to 10
	Kernel Function	linear, rbf
	Kernel Coefficient (γ)	0.1, 0.5, 1.0
PLS	Number of Components	1 to 10
Lasso	Regularization Parameter (α)	0.001 to 10

表 3.3: 深層学習におけるハイパーパラメータの探索範囲

Parameter Category	Parameter	Value
Basic Settings	Learning Rate	1.0×10^{-5} to 1.0×10^{-2}
	Activation Function	ReLU, Tanh, ELU, Swish
	Dropout Rate	0.1 to 0.5
Descriptor Network	Number of Layers	1 to 3
	Hidden Layer Units	10 to 190
Distribution Network	Number of Layers	1 to 3
	Hidden Layer Units	10 to 190
Concatenated Network	Number of Layers	1 to 3
	Hidden Layer Units	10 to 190
Training Settings	Batch Size	8, 16, 32
	Number of Epochs	50 to 300

3.3.5.2 精度指標

評価指標として、決定係数 (R^2) と二乗平均平方根誤差 (Root Mean Squared Error; RMSE) を用いた。 R^2 は予測値と実測値の相関の強さを評価する指標であり、RMSE は予測誤差の絶対的な大きさを評価する指標である。これらは以下の式で定義される。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.2)$$

ここで、 y_i は実測値、 $\hat{y}_i$ は予測値、 $\bar{y}$ は実測値の平均値、 n はサンプル数を表す。

3.3.5.3 効果量の統計解析

本研究では、Test データにおける RMSE を指標とし、3.3.2 で述べた記述子による特徴量拡張の有効性と再現性を検証した。まず、同一条件（学習モデルとデータ分割が同一）で得られたベースライン（MP）で得られた RMSE と記述子を追加した説明変数（MPD もしくは MPDD）で得られた RMSE との差分

$$d = \text{RMSE}(\text{MP}) - \text{RMSE}(\text{MPD もしくは MPDD})$$

を算出した。ここで $d > 0$ は拡張による RMSE 低下（予測精度向上）を意味する。

効果量の要約として、各特徴量集合ごとに、学習モデル別の平均改善量 $\bar{d}$ とその 95% ブートストラップ信頼区間 (CI) を算出した。ブートストラップの再標本化単位は（同一モデル、同一シード）の対応ペアであり、そのペアを単位とする復元抽出ブートストラップ（百分位法、繰返し数 $B = 50,000$ ）を用い、 $\text{CI}_{95\%} = (\text{CI}_{\text{low}}, \text{CI}_{\text{high}})$ を再抽出して平均差の CI を得た。加えて、検定に基づく補足評価として、各 $\{d\}$ に対する Wilcoxon 符号付き順位検定（両側、 $p < 0.05$ ）を行った。ただ、本検定は探索的分析と位置づけ、 p 値の多重比較補正は行わなかった。

また、上記の効果量要約（平均改善量）に対する再現性評価のため、50 個の乱数シード（3.3.5.1 で述べた 50 回）を「探索 (discovery)」25 個と「検証 (validation)」25 個に非重複で分割した。探索集合では、各（ターゲット目的変数、特徴量集合、学習モデル）について乱数シード 25 個に対する平均改善量 $\bar{d}^{(\text{disc})}$ を計算し、その後、同一の組に対して検証集合から得られた平均改善量 $\bar{d}^{(\text{valid})}$ を求めた。次に、 $\text{sign}(\bar{d}^{(\text{disc})}) = \text{sign}(\bar{d}^{(\text{valid})})$ のときに「方向一致」と判定した。方向一致率は、（ターゲット目的変数、特徴量集合、学習モデル）の組み合わせ数 n を分母とし、一致数 k との比 $r = k/n$ を算出した。本研究の設定では $n = 60$ （5 目的変数 $\times$ 2 特徴量集合 $\times$ 6 学習モデル）である。また、信頼区間の推定として、Clopper-Pearson による 95% 信頼区間を用いた。

3.3.6 外挿予測における評価設計

3.3.6.1 データ分割

外挿予測は、新規原料の採用や改善処方探索など実務上不可避な外挿条件下での妥当性を検証するために実施した。前提として、本研究における処方データは時系列順に並べられているが、製品リニューアル、もしくは新商品開発などによって新規原料や過去データセットにない原料組み合わせが発生する。そのため、過去データを学習データとして将来データを予測することで、外挿予測評価できると考えた。

外挿データにおけるデータ分割は、Rolling-origin に基づき、Train 窓を開始点固定で時系列方向に累積的に拡大する expanding window として設定し、その直後の検証 (Dev) とテスト (Test) を固定幅で評価する枠組みとすることで、データ分割による精度依存性を下げる手法を選択した。また、初期分割として Train:50%, Dev:20%, Test:10% を設定し、残り 20% を 4 分割した移動ステップ幅 5% で前進させることで、fold 数は 5 とした。この設計は、比率固定の単純な時系列分割 (例: 60/20/20) ではなく、初期窓と移動ステップを定めた Rolling-origin (expanding window) に基づく。

3.3.6.2 適用範囲 (Applicability Domain; AD) の定義

外挿予測においては、Rolling-origin によって設定した Test データが外挿となっているか定量的に検証するため、以下の 4 種類の適用範囲 (Applicability Domain; AD) 指標を採用した。

- レバレッジ (Leverage) : 線形モデルの Hat 行列に基づき、学習に用いるデータ (Train+Dev) に対するテストデータの外挿度を表す。Train+Dev における Hat 値の 95 パーセンタイル閾値を外挿判定の基準とした。
- マハラノビス距離 (Mahalanobis Distance; MD) : PCA により有効次元を推定し、平方距離に対するカイ二乗分布に基づく閾値で外挿を判定した。
- k 近傍平均距離 (kNN) : Train+Dev の k 近傍までの平均距離の 95 パーセンタイルを閾値として外挿を判定した。
- 学習範囲逸脱率 (Range) : 各特徴量の学習時の最小-最大範囲からの逸脱割合を算出した。

3.3.6.3 精度指標

内挿評価においては、評価指標として、決定係数 (R^2) と二乗平均平方根誤差 (RMSE) を利用したが、外挿評価においては、順位相関 (Spearman の ρ) も加えて評価する。これは、以下の式で定義する。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$
(3.3)

$$\rho = \text{corr}(\text{rank}(y), \text{rank}(\hat{y}))$$
(3.4)

ここで、 y_i は実測値、 $\hat{y}_i$ は予測値、 $\bar{y}$ は実測値の平均、 n はサンプル数を表す。また、 $\text{rank}(y)$ および $\text{rank}(\hat{y})$ は、それぞれ実測値ベクトル $y = (y_1, \dots, y_n)$ と予測値ベクトル $\hat{y} = (\hat{y}_1, \dots, \hat{y}_n)$ に付与した順位 (同順位は平均順位) を表し、 $\text{corr}(\cdot, \cdot)$ はピアソン相関係数を表す。等順位 (tie) が存在しない場合には、 $d_i = \text{rank}(y_i) - \text{rank}(\hat{y}_i)$ とおくと以下の式 (3.5) が成り立つ。

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$
(3.5)

3.4 結果と考察

3.4.1 内挿予測におけるモデル性能

Test データにおける性能比較 (R^2 と RMSE) を二軸で図 3.3～図 3.7 に示す. なお, 各図はベースラインである MP(Material + Process) とスカラー記述子群を追加した説明変数セット (MPD:MP + Descriptors, MPDD:MPD + Distribution) を比較可能な体裁で提示している (凡例は図 3.8 参照). また, 各目的変数における最良モデルの予測値と実測値の対応関係を散布図として図 3.9～図 3.13 に示す (50 回の乱数計算のうちランダムに 1 つの乱数シードを選択して Train+Dev と Test に関する散布図を描画した). さらに, MP を基準とした RMSE 改善量について百分位法ブートストラップ CI と Wilcoxon 符号付き順位検定に基づいて有意性を評価し, 結果をフォレストプロットとして図 3.14 に示す. 最後に, 効果の解釈として, MP を基準とした精度改善量の要約を表 3.4 に, また各目的変数における最良モデルでの精度比較を表 3.5 に示す.

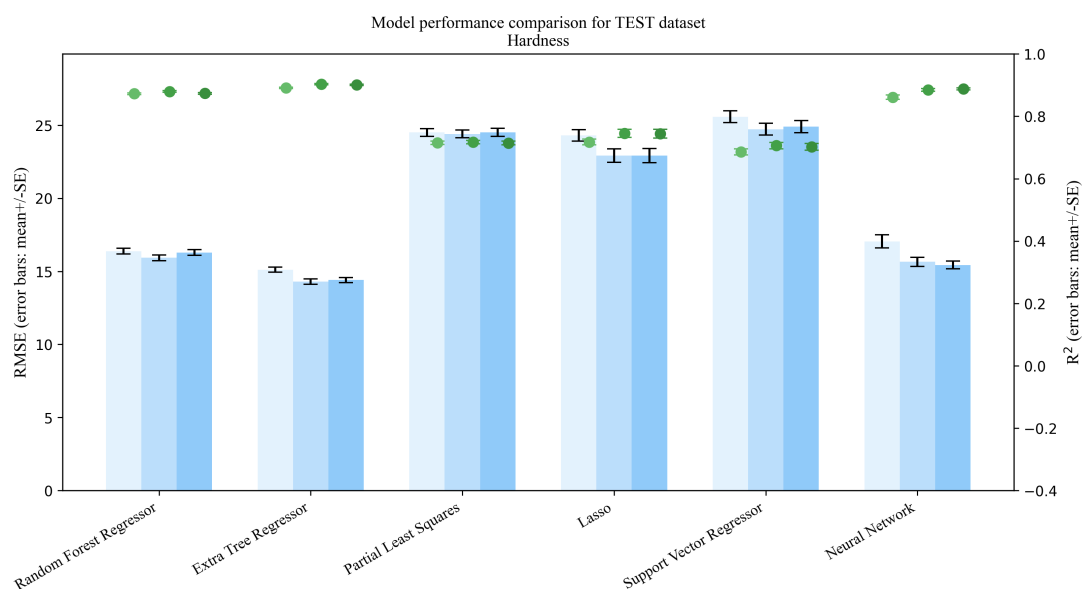


図 3.3: テストデータセットにおけるモデル性能比較 (硬度, $n = 1209$)

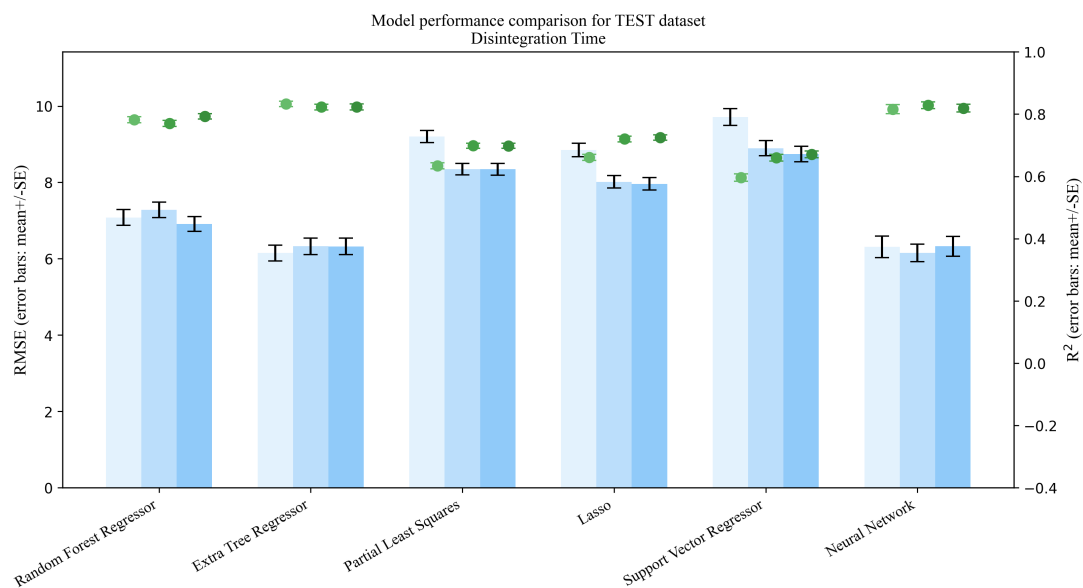


図 3.4: テストデータセットにおけるモデル性能比較（崩壊時間, $n = 882$ ）

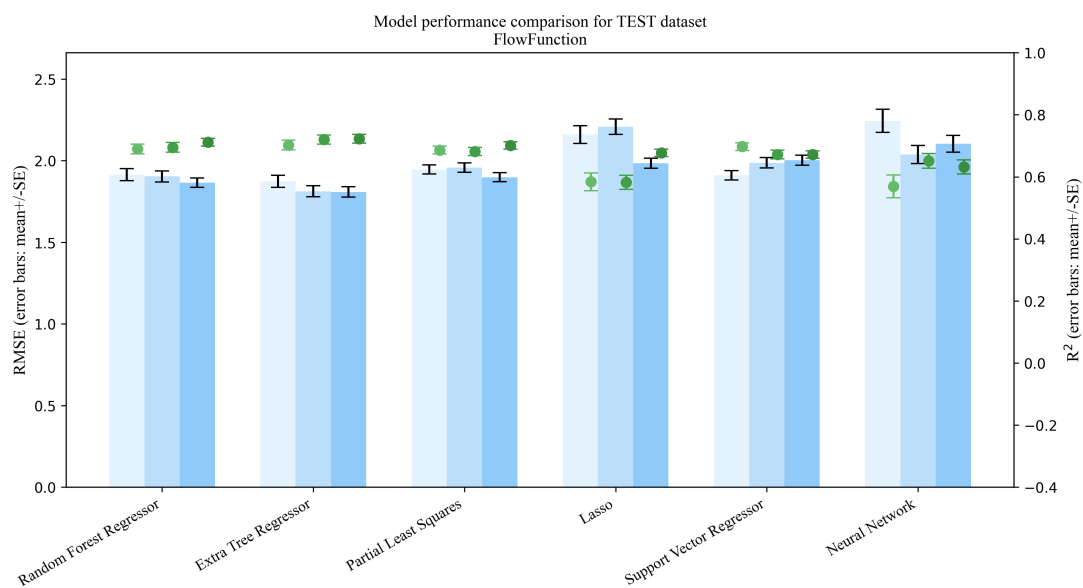


図 3.5: テストデータセットにおけるモデル性能比較（流動性指標, $n = 151$ ）

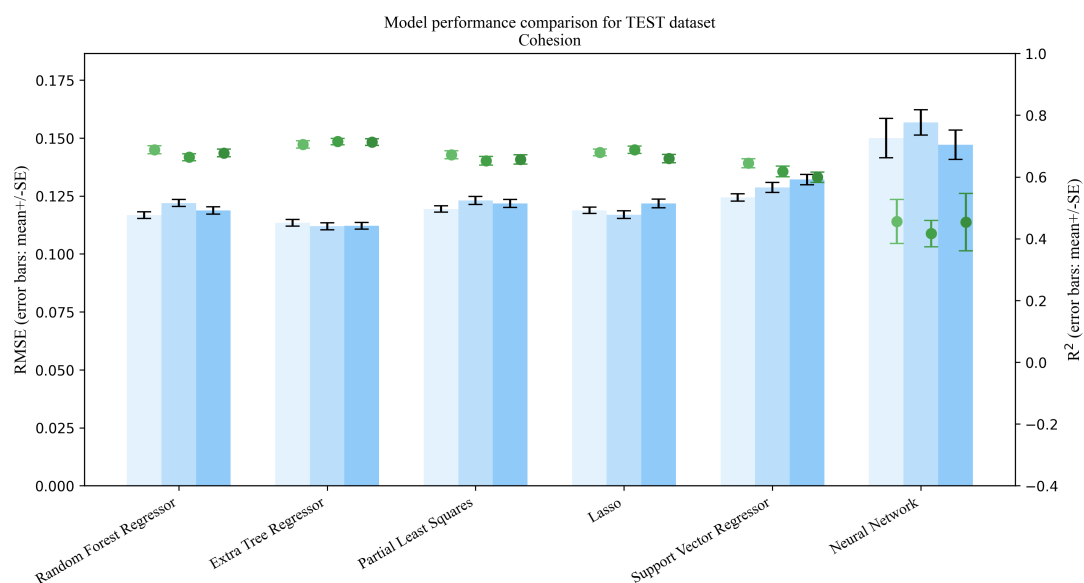


図 3.6: テストデータセットにおけるモデル性能比較（凝集性, $n = 145$ ）

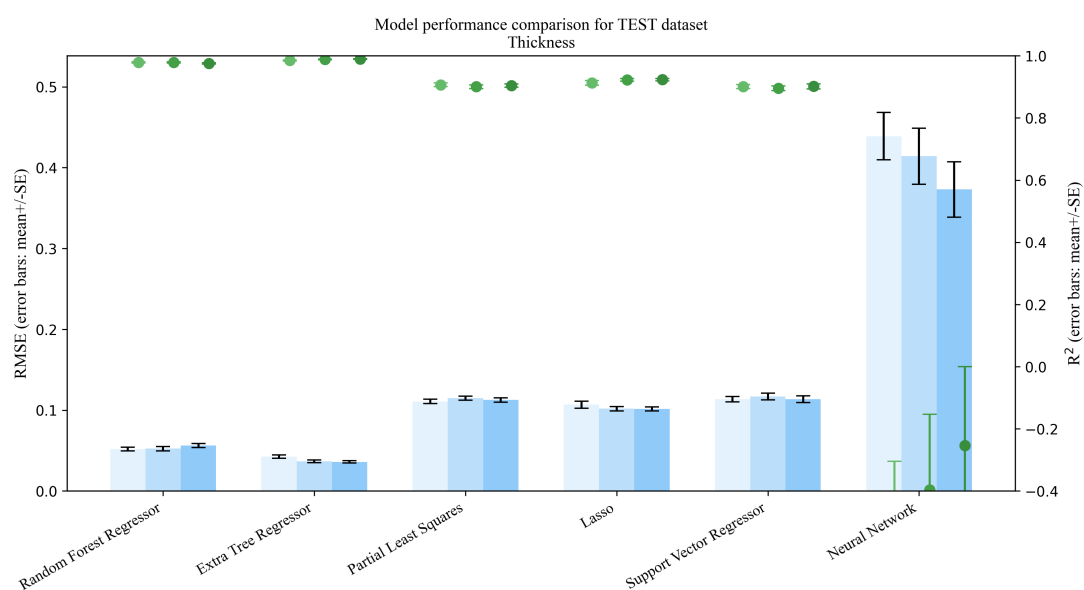


図 3.7: テストデータセットにおけるモデル性能比較（厚み, $n = 60$ ）

- RMSE by Materials and Processes (MP)
- RMSE by Materials, Processes, and Scalar descriptors (MPD)
- RMSE by Materials, Processes, Scalar descriptors, and Distributions (MPDD)
- R^2 by Materials and Processes (MP)
- R^2 by Materials, Processes, and Scalar descriptors (MPD)
- R^2 by Materials, Processes, Scalar descriptors, and Distributions (MPDD)

図 3.8: データパターンの凡例

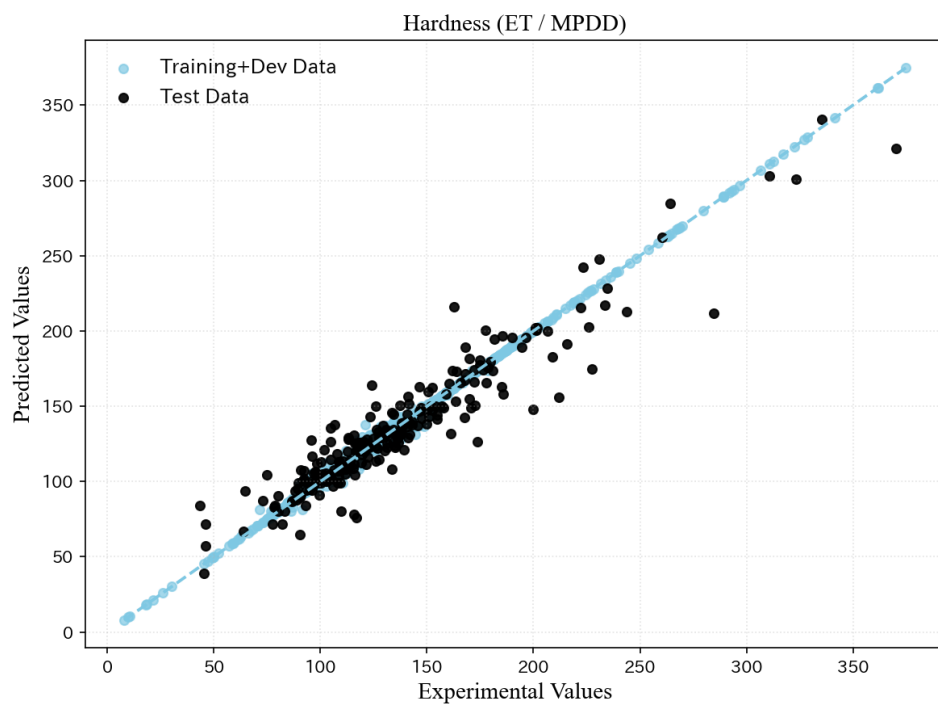


図 3.9: テストデータセットにおける最良モデルの散布図（硬度，ET，MPDD）

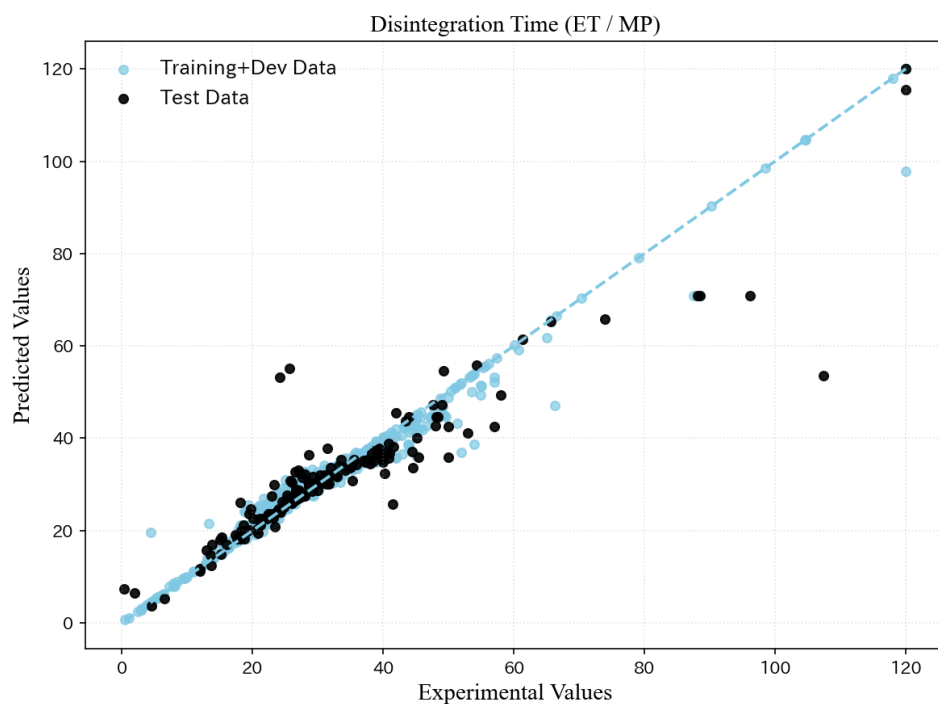


図 3.10: テストデータセットにおける最良モデルの散布図（崩壊時間，ET，MP）

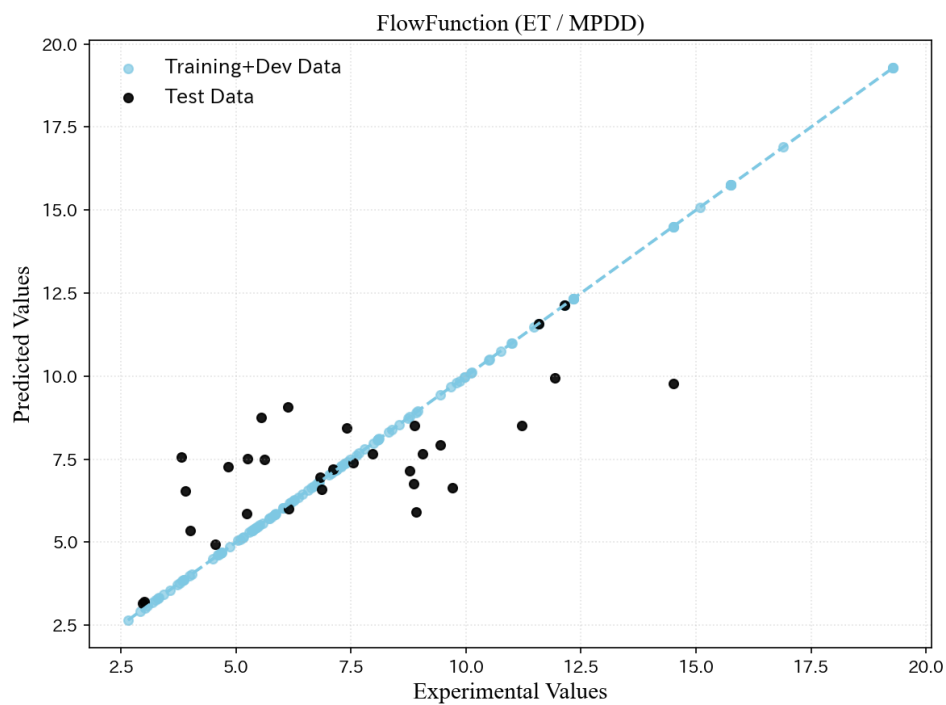


図 3.11: テストデータセットにおける最良モデルの散布図（流動性指標, ET, MPDD）

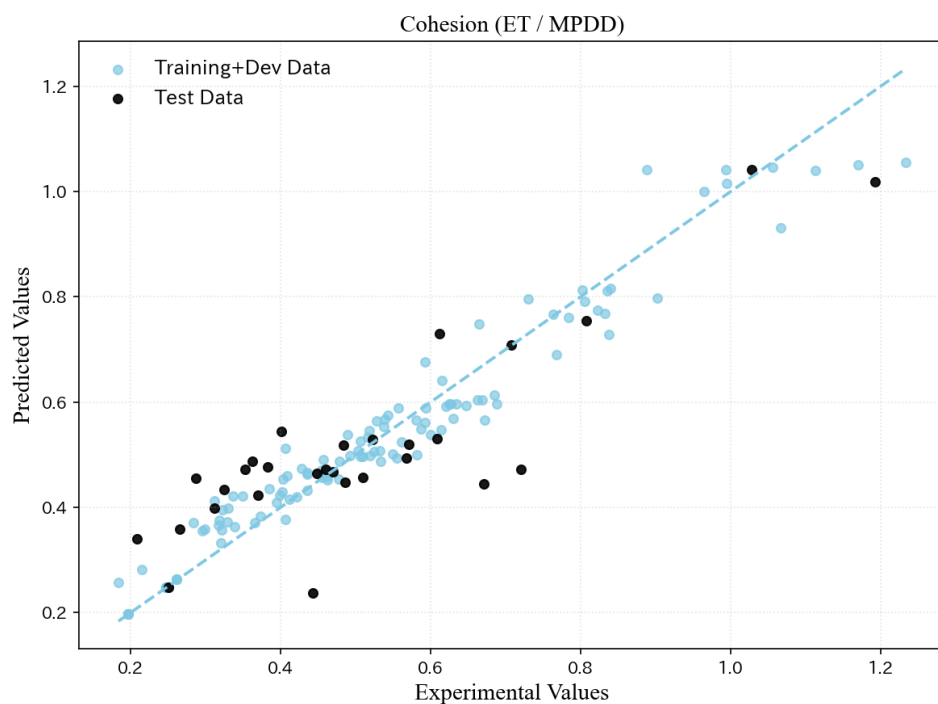


図 3.12: テストデータセットにおける最良モデルの散布図（凝集性, ET, MPDD）

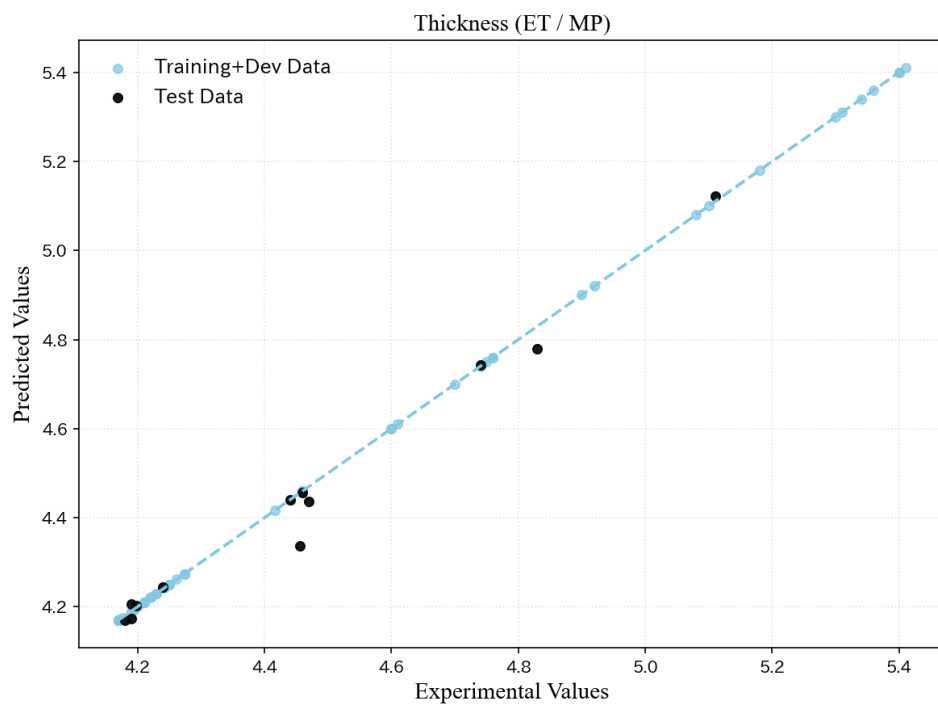


図 3.13: テストデータセットにおける最良モデルの散布図（厚み, ET, MP）

なお, MP を基準とした RMSE 改善量 $d = \text{RMSE}(\text{MP}) - \text{RMSE}(\text{MPD})$ もしくは $\text{RMSE}(\text{MPDD})$ を推定し, 百分位法ブートストラップ CI ($B = 50,000$) と Wilcoxon 符号付き順位検定 (両側) に基づいて有意性を評価した. その結果をフォレストプロットとして図 3.14 に示す.

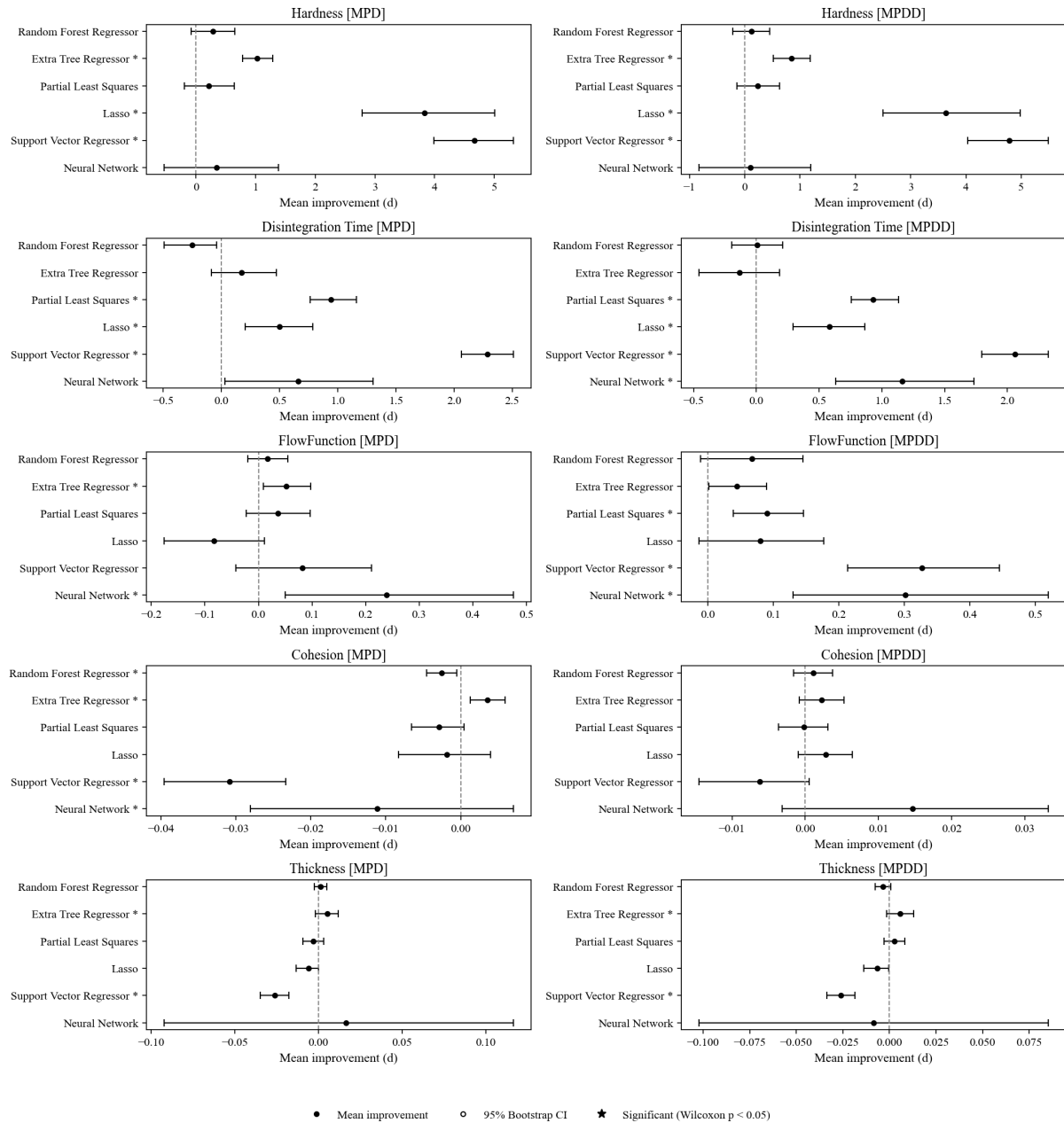


図 3.14: RMSE 改善量 $d = \text{RMSE}(\text{MP}) - \text{RMSE}(\text{MPDもしくはMPDD})$ のフォレストプロット. 各パネルはターゲット × 特徴量集合 (MPD, MPDD) を示し, 点は平均改善量 $\bar{d}$, 横線は 95% ブートストラップ CI (百分位法, $B = 50,000$). 縦破線は $d = 0$ (改善なし). “*” は Wilcoxon 符号付き順位検定 (両側) で $p < 0.05$ を示す. なお, 各目的変数のデータ数は, 硬度 $n = 1209$, 崩壊時間 $n = 882$, 流動性指標 $n = 151$, 凝集性 $n = 145$, 厚み $n = 60$ である.

総合的には, 硬度 (Hardness) では MPD/MPDD とともに平均改善量 $\bar{d}$ が正となるモデルが多く, 複数のモデルで 95% ブートストラップ CI が 0 を跨がず, Wilcoxon 符号付き順位検定 (両側) でも有意であった. 崩壊時間 (Disintegration Time) では, MPD は PLS/Lasso/SVR で改善が有意である一方, RF では平均改善量が負となり, モデル依存性が残る. MPDD では Lasso/NN/PLS/SVR で改善が有意である一方, RF は改善がほぼゼロ, ET では平均改善量がわずかに負となるなど, 改善量のばらつきが見られる. 流動性指標 (Flow Function) では効果量は小さいものの, MPDD では NN/PLS/SVR で正方向

の改善が有意となり，悪化が有意となるモデルは確認されなかった．凝集性 (Cohesion) では MPD で改善と悪化が混在し，RF/NN/SVR では悪化が有意となる．一方 MPDD では有意な改善・悪化はいずれも確認されず，効果の不確実性が残る．厚み (Thickness) はサンプル数が小さく推定が不安定であり，改善と悪化の双方が観察されるため，効果の一般化には留意が必要である．

また，探索 (discovery) セットにおいても，検証 (validation) セットにおいても $\bar{d}$ の符号が一致したのは $k = 50$ 組 (全 $n = 60$ 組) であり，方向一致率は $r = 0.833$ (Clopper–Pearson の 95%CI: 0.719–0.906) であった． r が 0.833 であり，95%CI の下限値も 0.5 を上回っていることから，方向一致率の観点では再現性に大きな問題はないと考える．

最後に，MP を基準とした精度改善量の要約を表 3.4 に，また各目的変数における最良モデルでの精度比較を表 3.5 に示す．ここで，表 3.4 は学習モデル群 (6 種) を集約した要約であり，MP の精度は各目的変数における全モデルの MP 設定での精度 (平均) を示す．一方，表 3.5 は各目的変数における最良モデルに限定した指標の平均であるため，両者の数値は一般に一致しない．表 3.4 より，硬度 (Hardness)・崩壊時間 (Disintegration Time)・流動性指標 (Flow Function) では MPD・MPDD とともに平均的に改善が確認され，とくに流動性指標 (Flow Function) は MPDD で改善量が多い．一方，凝集性 (Cohesion) では MPD が平均的に悪化するのに対し MPDD はわずかな改善に留まり，厚み (Thickness) では MPD/MPDD とともに改善は確認されない．また，最良モデルに限定すると (表 3.5)，改善量の集約とは異なる結論となる目的変数もあるため (例：崩壊時間 (Disintegration Time) で MPDD が MP を下回る)，目的変数ごとの運用では両表を併せて解釈する必要がある．

表 3.4: 内挿設定 (Test) における MP→MPD/MPDD の精度改善量の要約 (目的変数別, 平均)

Target	指標	MP の精度	MP→MPD の改善量	MP→MPDD の改善量
硬度 (Hardness)	R^2	0.768	0.040 ↑	0.039 ↑
	RMSE	21.602	1.728 ↑	1.621 ↑
崩壊時間 (Disintegration Time)	R^2	0.657	0.059 ↑	0.062 ↑
	RMSE	8.874	0.719 ↑	0.768 ↑
流動性指標 (Flow Function)	R^2	0.591	0.029 ↑	0.064 ↑
	RMSE	2.141	0.057 ↑	0.152 ↑
凝集性 (Cohesion)	R^2	0.635	−0.047 ↓	0.013 ↑
	RMSE	0.131	−0.008 ↓	0.002 ↑
厚み (Thickness)	R^2	0.737	−0.041 ↓	−0.041 ↓
	RMSE	0.129	−0.002 ↓	−0.006 ↓

注：改善量は， R^2 は $R^2(\text{MPD/MPDD}) - R^2(\text{MP})$ ，RMSE は $\text{RMSE}(\text{MP}) - \text{RMSE}(\text{MPD/MPDD})$ とし，正の値が改善となるように統一した．表中の矢印は改善量の解釈を示し，↑ は改善，↓ は悪化を表す．MP の精度は，各目的変数における全モデルの MP 設定での平均精度を示す．改善量は，学習モデル 6 種 × 乱数分割反復実験 ($n = 50$) の全試行に対する平均を示す．

表 3.5: 内挿設定 (Test) における最良モデルでの精度比較 (目的変数別, 平均)

Target	指標	MP	MPD	MPDD
硬度 (Hardness)	R^2	0.887	0.901	0.899
	RMSE	15.644	14.615	14.795
崩壊時間 (Disintegration Time)	R^2	0.809	0.819	0.801
	RMSE	6.661	6.487	6.796
流動性指標 (Flow Function)	R^2	0.687	0.700	0.700
	RMSE	1.913	1.861	1.868
凝集性 (Cohesion)	R^2	0.726	0.741	0.733
	RMSE	0.116	0.113	0.114
厚み (Thickness)	R^2	0.980	0.983	0.983
	RMSE	0.044	0.039	0.038

3.4.2 外挿予測におけるモデル性能

外挿評価を目的とした Rolling-origin 時系列分割における AD カバレッジの比較を表 3.6 に示す. 各目的変数 $\times$ 特徴量集合 (MP/MPD/MPDD) について, Leverage (Lev) $\cdot$ Mahalanobis 距離 (MD) $\cdot$ kNN $\cdot$ 学習範囲 (各特徴量の最小-最大) 内に収まる割合 (Range OK) の Train+Dev/Test fold 平均を要約した.

表 3.6 より, Train+Dev 側の AD カバレッジはいずれの指標でも概ね高水準である. 一方, Test 側では Leverage が低下し (例: 硬度 (Hardness)/MP で Lev: 43.8%, 硬度 (Hardness)/MPD で Lev: 23.1%, 崩壊時間 (Disintegration Time)/MPD で Lev: 40.2%), 将来側データが学習空間の外縁に位置する割合が一定程度あることが示唆される. このとき, MD や Range OK は Test 側でも高水準に留まるため, AD 指標間で外挿性の検出感度が異なることが確認できる.

表 3.6: Rolling-origin 時系列 CV における AD カバレッジ (fold 平均, Train+Dev/Test)

Target	Pattern	Split	Lev (%)	MD (%)	kNN (%)	Range OK (%)
硬度 (Hardness)	MP	Train+Dev	94.9	92.2	97.3	100.0
		Test	43.8	100.0	81.5	99.7
	MPD	Train+Dev	94.9	87.3	97.3	100.0
		Test	23.1	93.6	75.2	99.7
	MPDD	Train+Dev	94.9	86.2	97.1	100.0
		Test	21.8	93.9	75.5	99.7
崩壊時間 (Disintegration Time)	MP	Train+Dev	95.0	96.3	97.6	100.0
		Test	67.5	100.0	86.1	99.8
	MPD	Train+Dev	95.0	89.1	96.3	100.0
		Test	40.2	92.3	80.9	99.7
	MPDD	Train+Dev	95.0	88.5	97.1	100.0
		Test	37.5	92.7	80.9	99.7

上記を踏まえ, MP をベースラインとして, 記述子 (MPD) および粒度分布 (MPDD) の追加による精度変化を「精度改善量」として要約する. ここで, R^2 および Spearman の ρ の改善量は $R^2(\text{MPD/MPDD}) - R^2(\text{MP})$, $\rho(\text{MPD/MPDD}) - \rho(\text{MP})$, RMSE の改善量は $\text{RMSE}(\text{MP}) - \text{RMSE}(\text{MPD/MPDD})$ とし, いずれも正の値が改善を意味する.

複数モデルを集約した要約として, NN を除く 5 モデル (RF/ET/Lasso/PLS/SVR) に対して改善量を算出し, 平均で要約した結果を表 3.7 に示す.

追加の補足評価として, MP における最良モデル (硬度 (Hardness): RF, 崩壊時間 (Disintegration Time): Lasso) を固定し, 同一モデルでの指標 (MP/MPD/MPDD) を比較した結果を表 3.8 に示す.

複数モデルを集約した要約と, 最良モデル固定の追加評価のいずれにおいても, 硬度 (Hardness) では R^2 改善量 > 0 , RMSE 改善量 > 0 , ρ 改善量 > 0 となり, 外挿条件でも

表 3.7: 硬度 (Hardness)/崩壊時間 (Disintegration Time) の外挿設定 (時系列 CV) における要約 (複数モデル集約): 精度 (MP) および精度改善量の要約 (NN を除く 5 モデル)

(a) 硬度 (Hardness) ($n_{\text{models}} = 5$) におけるベースライン (MP) 精度と改善量

指標	MP の精度	MP→MPD の精度改善量	MP→MPDD の精度改善量
R^2	-0.193	0.223 ↑	0.230 ↑
RMSE	20.381	1.827 ↑	1.602 ↑
ρ	0.659	0.040 ↑	0.032 ↑

(b) 崩壊時間 (Disintegration Time) ($n_{\text{models}} = 5$) におけるベースライン (MP) 精度と改善量

指標	MP の精度	MP→MPD の精度改善量	MP→MPDD の精度改善量
R^2	0.089	-0.604 ↓	-0.660 ↓
RMSE	9.537	-1.242 ↓	-1.583 ↓
ρ	0.639	-0.092 ↓	-0.181 ↓

注: 表中の矢印は改善量 (平均) の解釈を示し, ↑ は改善, ↓ は悪化を表す. 改善量は, R^2 と ρ は $R^2(\text{MPD/MPDD}) - R^2(\text{MP})$, $\rho(\text{MPD/MPDD}) - \rho(\text{MP})$, RMSE は $\text{RMSE}(\text{MP}) - \text{RMSE}(\text{MPD/MPDD})$ として算出し, 正の値が改善となるように統一した.

表 3.8: 硬度 (Hardness)/崩壊時間 (Disintegration Time) の外挿設定 (時系列 CV) における追加の補足評価: MP で最良モデルを固定した指標比較 (Test 平均)

(a) 硬度 (Hardness) における最大精度モデル (RF) の精度比較

指標	MP における精度	MPD における精度	MPDD における精度
R^2	0.355	0.520	0.494
RMSE	16.320	13.926	14.418
ρ	0.738	0.811	0.815

(b) 崩壊時間 (Disintegration Time) における最大精度モデル (Lasso) の精度比較

指標	MP における精度	MPD における精度	MPDD における精度
R^2	0.408	-2.335	-2.031
RMSE	8.006	14.787	14.178
ρ	0.718	0.564	0.551

記述子/分布特徴の追加が改善傾向を示した。一方、崩壊時間 (Disintegration Time) では R^2 改善量と ρ 改善量が負となり、順位の単調性が低下する傾向が示された。また、RMSE 改善量も平均では負となり、総体として悪化傾向が確認された。

推定精度低下の原因としては、(i) 追加記述子や分布特徴が学習 fold と将来 fold の分布差 (共変量シフト) を増幅し、高次元化により過適合を助長すること、(ii) 記述子の一部が情報冗長・高相関で外挿境界におけるノイズを増幅すること、が考えられる。

3.4.3 物理的考察

本節では、各目的変数の支配機構と、追加記述子 (MPD: スカラー記述子を加えた特徴量, MPDD: MPD に粒度分布の分布ベクトルに基づく要約特徴量) がその機構をどの程度表現しうるかを踏まえて、精度変化の傾向を定性的に解釈する。括弧内は解析に用いたサンプル数である。

硬度 (Hardness, $n = 1209$): 錠剤強度は圧縮下での塑性変形/脆性破壊と粒子間接触・実効結合面積 (空隙率) の変化に支配される。密度、吸湿性、形状などのスカラー記述子 (MPD) はこれらの平均的挙動に強く対応するため、全モデル平均の指標でも MP 対比で改善が得られたのは妥当である。粒度分布 (MPDD) は本来、微粉の充填・架橋を通じて結合面積に影響するが、高次元かつ相関の強い分布特徴は効果量を減じたと考えられる。それでも平均指標および最良モデルのいずれでも MPDD での改善が確認され、結合面積に関わる情報の付与が有効であることを示唆する。

崩壊時間 (Disintegration Time, $n = 882$): 崩壊は毛管浸透と崩壊剤の膨潤に支配され、細孔網構造 (空隙率・細孔径分布) が鍵となる。平均粒径や濡れ性などの MPD は平均的な浸潤挙動を捉え、粒度分布 (MPDD) は細孔径分布の代理として機能するため説明力が増す。平均指標で見ると、内挿評価において MPD, MPDD 共に上乘せ効果が見られた。

流動性指標 (Flow Function, $n = 151$): 流動性は粒子間摩擦、形状異方性、微粉凝集、水分の微小変動に敏感で、実験再現性の確保が難しい指標である。MPD の形状・表面特性は寄与するが、信号に対するノイズが相対的に大きく、平均的な効果量は小さくなりやすいと考えられる。MPDD に含まれる微粉端の情報は有益であり、平均指標および最良モデルの双方で正方向の改善が確認された。

凝集性 (Cohesion, $n = 145$): 凝集は微粉分率、表面エネルギー、湿潤架橋に強く依存し、粒度分布は一次的決定因子である。一方で、本研究の内挿評価では効果量自体が小さく、平均指標で見ても改善は限定的であった。MPD では平均的にわずかな悪化が観測され、最良モデルにおいても上乘せ効果は明確ではなかった。一方で MPDD では有意な改善・悪化はいずれも確認されず、上乘せ効果は不確実であった。データ数 $n = 145$ が比較的小さく推定のばらつきが大きいことに加え、凝集性に関わる微粉端・表面状態などの情報が、分布特徴だけでは十分に表現できていない可能性もある。

厚み (Thickness, $n = 60$): 厚みは充填密度と圧縮応答 (弾性回復) で決まり、粒度分布は空隙充填とパッキングを通じて本来有効であると考えられる。原理的に MPDD が有効と考えられるが、本研究では平均指標でも最良モデルでも効果が一貫しなかった。主因として、 $n = 60$ と小さく高次元分布特徴に対して自由度不足であること、また、NN 以外のモデルでは R^2 が 0.9 以上に達しており、さらなる精度向上の余地が限定的であった可能性がある。

3.5 結論

本章では、錠剤開発における配合設計・工程設定の最適化を支援するため、材料・プロセス情報 (MP) にスカラー記述子 (D) とさらに粒度分布 (DD) を統合した特徴量 (MP/

MPD / MPDD) と機械学習・深層学習から成る予測モデル群を構築した。混合系は配合比による重み付き和で記述子を定式化した。Train / Dev / Test 分割と反復実験 ($n = 50$) により再現性を担保し、RMSE 改善量 d のブートストラップ CI と Wilcoxon 検定により効果を評価した。

内挿空間予測における Test 結果 (図 3.3～図 3.7) において精度比較を提供し、効果量要約のフォレストプロット (図 3.14) と整合して、硬度および崩壊時間を中心に特徴量拡張 (特に MPD) の有効性が平均改善量ベースで確認された。組合せにより悪化も一部生じたが、平均改善量で見ると硬度・崩壊時間・流動性指標では改善が比較的明確であった。また、MPDD による流動性分布情報の追加も平均的に精度向上に寄与した。

外挿性評価では、Rolling-origin 時系列分割により Train/Test 分布の乖離を確保し、厳しい外挿条件で検証した。その結果、硬度では MPD/MPDD とともに RMSE 改善量 d が正 (RMSE 低下) となり、 ΔR^2 および Spearman の ρ も正となったことから、記述子・分布特徴の追加は改善傾向を示した。一方、崩壊時間では ΔR^2 および $\Delta \rho$ が負であり、RMSE 改善量も負 (RMSE 増加) となるなど、総体として悪化傾向が確認された。また、AD が高い領域では RMSE が低下する傾向が確認され、高次元特徴の追加は外挿で過適合を招きうることから、記述子選択と次元抑制が重要である。

なお、本章の目的で掲げた課題に対する回答は次のとおりである。内挿では硬度および崩壊時間、流動性指標における平均改善量が正となった一方、凝集性と厚みでは平均改善量が小さく、上乗せ効果は限定的であった (凝集性の MPDD は平均としては改善方向だが効果量が小さい)。記述子拡張の効果はブートストラップ CI と Wilcoxon 検定により統計的に支持され、乱数分割の方向一致率は $r = 0.833$ ($k = 50/n = 60$, Clopper-Pearson の 95%CI: 0.719–0.906) で再現性が示された。また、追加特徴 (記述子・分布特徴) と支配機構の対応を定性的に整理することで、モデル出力の解釈可能性に関する示唆を得た。外挿評価では Rolling-origin と AD により外挿性を担保し、硬度では記述子・分布特徴の追加が改善傾向を示した一方、崩壊時間では指標全体として悪化傾向が確認された。AD カバレッジが高い領域ほど誤差が低い傾向は維持され、高次元特徴の追加における次元抑制の重要性が示唆された。

総括すると、材料・プロセス・記述子・粒度分布の統合と多様な学習モデルの併用により、錠剤物性の予測精度を実務上有用な水準へ高めうることを示した。本枠組みは、配合設計の探索効率化や知見の形式知化を通じて、開発期間短縮・試験コスト削減・品質安定化に寄与しうる。

Chapter 4

クレンジング剤開発への AI の活用

4.1 本章の目的

本研究で対象とするクレンジング剤は水性クレンジング（洗顔フォーム）である。以降、本章では洗顔フォーム（クレンジング剤）を単にクレンジング剤と表記する。本章では、化粧品業界における処方設計の課題解決を目指し、人工知能（AI）技術を活用したクレンジング剤の性能予測システムの構築について述べる。クレンジング剤は多数の成分で構成されており、成分間の相互作用や自己組織化現象が洗浄性能に大きく影響するため、従来の経験的アプローチでは最適な処方を見つけることが困難である。本研究では、分子記述子やハンセン溶解度パラメータなどの化学的特性指標を用いて成分の物理化学的性質を数値化し、機械学習による非線形関係のモデル化により、成分間の複雑な相互作用を考慮した洗浄能力予測システムを構築する。さらに、逆解析を導入することで、実際の実験を行うことなく大量の処方候補を評価し、開発期間の短縮とコスト削減を実現する。このアプローチにより、従来の経験則に頼らない科学的根拠に基づく処方設計手法の確立を目指す。

4.2 実験手法

4.2.1 洗浄率の評価

クレンジング剤はイオン性界面活性剤、両性界面活性剤、ノニオン界面活性剤、ポリオール、pH 調整剤、水で構成される（使用原料例を表 4.1 に示す）。これらを、加熱しながら十分に混合・攪拌することにより調製した。調製したクレンジング剤のサンプル例を表 4.2 に示す。

本章の目的変数となる洗浄率の測定手順の概略図を図 4.1 に示す。手法として、まず白色の人工皮革に防水性ペンシルアイライナーを塗布し、30 分間乾燥させた。その後、対応するクレンジング剤サンプル 0.1 mL を乾燥したペンシルアイライナー上に添加し、30 回擦った後、すすいで乾燥させた。洗浄率は、以下の式により色差から算出したペンシルアイライナーの残留率を色差にて評価した [15]：

$$\text{洗浄率(\%)} = \frac{\sqrt{(L_0^* - L_1^*)^2 + (a_0^* - a_1^*)^2 + (b_0^* - b_1^*)^2} - \sqrt{(L_0^* - L_2^*)^2 + (a_0^* - a_2^*)^2 + (b_0^* - b_2^*)^2}}{\sqrt{(L_0^* - L_1^*)^2 + (a_0^* - a_1^*)^2 + (b_0^* - b_1^*)^2}} \times 100 \quad (4.1)$$

表 4.1: 本研究で使用した原料カテゴリーの例

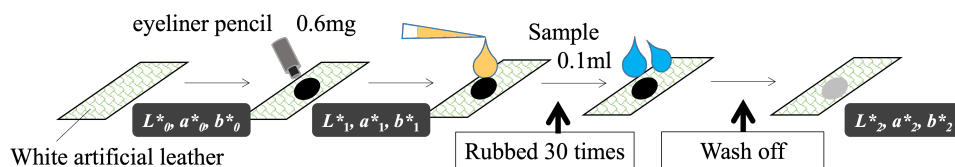
Category	The Number of Ingredients	Ingredient Examples
Anionic surfactant	8	Potassium Cocoyl Glutamate; Potassium Cocoyl Glycinate
Amphoteric surfactant	4	Lauramidopropyl Hydroxysultaine; Sodium Cocoamphoacetate
Nonionic surfactant	24	Eicosaglycerol Hexacaprylate; Decaglycerol Isostearate; PEG-20 Glyceryl Triisostearate
Polyol	33	PPG-9 Diglyceryl Ether; Cyclohexylglycerin; Glycerin
pH adjuster	1	Citric Acid
Base	1	Water

表 4.2: 本研究におけるクレンジング剤サンプル例（実験例；組成は wt%）

Category	Material Name	Ex.A	Ex.B	Ex.C	Ex.D	Ex.E	Ex.F
Anionic surfactant	Potassium Cocoyl Glutamate	6	7	6	7	13	9
Amphoteric surfactant	Lauramidopropyl Hydroxysultaine	5	9	7	8	4	6
Nonionic surfactants	Eicosaglycerol Hexacaprylate	10	10	0	0	0	0
	Decaglycerol Isostearate	0	0	0	0	10	0
	PEG-20 Glyceryl Triisostearate	0	0	7	0	0	0
Polyols	PPG-9 Diglyceryl Ether	0	11	0	0	0	14
	Cyclohexylglycerin	9	0	6	11		0
	Glycerin	0	0	0	0	14	0
pH adjuster	Citric Acid	0.8	0.8	0.8	0.8	0.8	0.8
Base	Water	69.2	62.2	73.2	73.2	58.2	70.2

ここで、 L^* は明度、 a^* および b^* は色度を表し、 $L^*a^*b^*$ は色彩計 (CM-2600d, Konica Minolta, Inc.) で測定した色空間である. (L_0^*, a_0^*, b_0^*) , (L_1^*, a_1^*, b_1^*) , (L_2^*, a_2^*, b_2^*) は、それぞれペンシルアイライナー塗布前、塗布後、クレンジング後の白色人工皮革の色空間値を表す [15].

Evaluation of cleansing capability



L^* , a^* , b^* value were measured by a colorimeter (CM-2600d, Konica Minolta, Inc.)

図 4.1: 本研究で調製した処方の洗浄率を決定する評価テストの概要

4.2.2 AIモデリング手法

データ処理フローチャートを図 4.2 に示す.

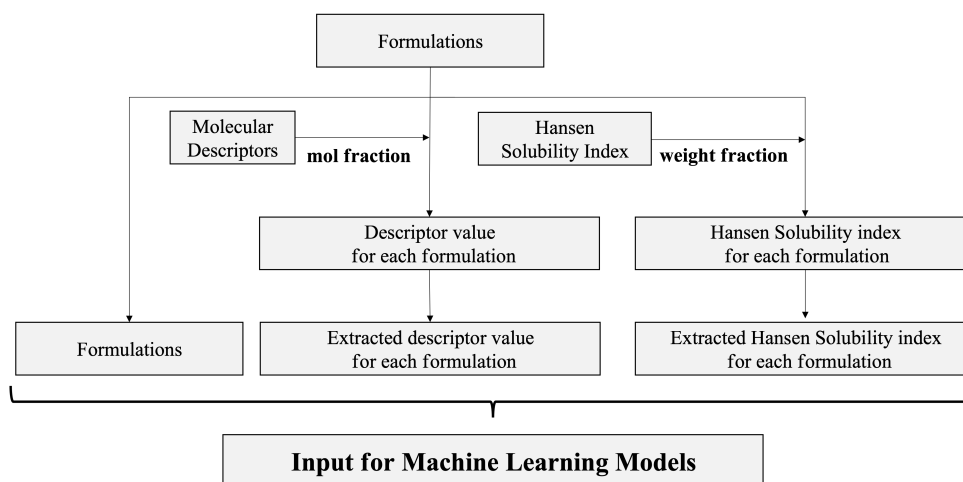


図 4.2: データ処理フロー

4.2.2.1 分子記述子

分子記述子は、化学構造から抽出される数値的な基本分子特性として定義される. 各種分子記述子は、特定の分子内の化学基間の特定のタイプの相互作用に関連している. 記述子は、単一化学物質だけでなく、化学混合物の化学的性質の予測にも利用される. また、記述子はジェミニ界面活性剤の CMC 予測にも応用されている [33]. そこで、原料から情報を抽出し、分子記述子を用いて調製した処方の洗浄率を予測した.

これらの値は、ケモインフォマティクスツールである RDKit と PaDEL-descriptor を用いて、各原料の化学構造式から算出した [52, 53]. 各原料の構造式は ChemDraw で作成し、SMILES へ変換した. 重合度を有する原料については、各原料に対して代表的な重

合度を採用した．無限大値や単一値の特徴量は除去し，標準化した後， k -近傍法補完器 ($k = 5$) を用いて欠損値を予測した．その後，混合原料の記述子値を推定するために，モル分率または重量分率を用いて原料の加重平均を算出した．

4.2.2.2 ハンセン溶解度パラメータ (HSPs)

洗浄率の予測にハンセン溶解度パラメータ (HSPs) を適用した．HSPs は，Charles M. Hansen によって開発され，物質が別の物質に溶解して溶液を形成する能力を予測するものである．HSPs は通常，溶質と溶媒の HSPs 距離を計算することにより，溶質が溶媒に溶解するかどうかを推定するために使用される．界面活性剤の性質予測に HSPs を利用した報告もいくつかある [54, 55]．

本研究における溶質であるアイライナーが多くの原料で構成されており，その構造式の特定が困難だったため，本研究では溶質と溶媒の距離算出の代替として，最高の洗浄率を有するクレンジング剤サンプルの処方と各サンプル処方との間の距離を計算した．

HSPs 距離は $\{4 * (dD_1 - dD_2)^2 + (dP_1 - dP_2)^2 + (dH_1 - dH_2)^2\}^{0.5}$ で定義され，ここで dD_1 , dP_1 , dH_1 は各サンプルの値（混合物については重量分率による加重平均）であり， dD_2 , dP_2 , dH_2 は最高洗浄率を示す 3 処方の HSPs 値の平均値である．HSPs は，原料間の相互作用の非線形効果を測定することが困難な記述子法よりも，処方中の原料間の相互作用の効果をより良く推定できると期待した．

HSPs 値は HSPiP ソフトウェア (version 5.0.09) によって計算した．高分子量分子については，HSPiP によって HSPs を計算できないものがあるため，記述子計算と同様に，欠損 HSPs 値は k -近傍法補完器 ($k = 5$) によって補完した．

4.2.2.3 モデリングと特徴選択

記述子と HSPs の使用により，説明変数の数は 1,000 を超えるため，これらの多数の特徴からクレンジング性能を予測する法則を得るために機械学習を適用した．機械学習アルゴリズムには，教師あり学習，教師なし学習，強化学習の 3 つのタイプがある．本研究の目的は連続値の出力を予測することであったため，教師あり学習（回帰）を選択した．

入力データセットは 4.2.2.1 節と 4.2.2.2 節で説明したものであり，出力は洗浄率である．また，全ての説明変数は平均 0，標準偏差 1 で正規化されたが，Test データのリークageを防ぐため，各交差検証 fold ごとに独立して正規化を実施した．正規化パラメータ（平均と標準偏差）は訓練データのみから計算され，各 fold 内の訓練データと Test データの両方に適用された．学習モデルとして，2 つの決定木ベースモデル（Random Forest Regressor と Extremely Randomized Trees (Extra-Trees, ET) Regressor），2 つの線形ベースモデル（Lasso と Partial Least Squares），1 つのサポートベクターベースモデル（SVR）を採用した．また，ハイパーパラメータを表 4.3 に示す．ハイパーパラメータはグリッドサーチ法に基づいて最適化した．

一方，多数の特徴量はモデリングにノイズを引き起こすため，重要な特徴量を抽出するために Boruta 法 [56] も採用した．

4.2.2.4 モデル評価

データセットは 537 サンプルで構成され，精度計算には 10 分割交差検証が適用された．計算ではデータの 90% を訓練データに割り当て，残りの 10% を Test データとした．この計算は 10 回実施され，全データがいずれかの時点で Test データとして使用されることを保証し， R^2 スコアの平均値をモデル性能として採用した．また，分割乱数の依存性を排除するため，異なる乱数シードを用いて 5 回実行し，分割乱数の偏りが無いことを確認している．

表 4.3: モデリングにおけるハイパーパラメータ設定

Model Name	Hyperparameter	Values
Extra-Trees	Number of Estimators	10, 50, 100, 500, 1000, 2000
	Maximum Features	(the number of features) ^{0.5} , $\log_2(\text{the number of features})$, the number of features
	Maximum Depth	10, 20, 30
Random forest regressor	Number of Estimators	10, 50, 100, 500, 1000, 2000
	Maximum Features	sqrt, log2, none
	Maximum Depth	10, 20, 30
Support vector regressor	Regularization Parameter (C)	0.1, 1, 10
	Epsilon	0.01, 0.1, 1
	Kernel Function	Linear, poly, rbf
Partial least squares regressor	Number of Components	1 to 20
Lasso regressor	Regularization Parameter (α)	0.0001, 0.001, 0.01, 0.1, 10

4.2.3 逆解析による処方設計と特徴量効果の可視化

どの原料，もしくは原料の組み合わせが洗浄率に寄与しているかを検証するために，逆解析によって仮想処方を作成しその洗浄率を予測した上で層別解析を行う．本研究ではノニオン界面活性剤とポリオールの効果検証を行う．このとき，処方は乱数を用いて作成するが，処方の配合比率には一定に制約条件があること，また，今回の観察したいノニオン界面活性剤とポリオールの効果の可視化に特化するため，以下で説明するルールに従って仮想処方を作成した．

- 全ての原料は6つのカテゴリー（表 4.1 と同じカテゴリー）のいずれかに分類された：アニオン性界面活性剤，両性界面活性剤，ノニオン界面活性剤，ポリオール，pH 調整剤（クエン酸のみ），ベース（水のみ）．
- ノニオン界面活性剤とポリオールに焦点を当てるため，アニオン性および両性界面活性剤の選択をそれぞれ1種類に制限した．このとき，データセットにおける利用頻度の多さからアニオン性界面活性剤は Potassium Cocoyl Glutamate を，両性界面活性剤は Lauramidopropyl Hydroxysultaine を選択した．
- ノニオン界面活性剤とポリオールそれぞれにおいて原料セットから1種類ずつ原料種を選択した．（そのため，例えば，1つの処方に2つのノニオン界面活性剤を選択することはできない）
- pH 調整剤（クエン酸）と水を除く各原料の添加率は，表 4.4 に記載された予め定義された範囲内で，各カテゴリーについてランダム化した．クエン酸の添加率は 0.8 wt% で固定し，水の添加率は全原料の合計が 100 wt% になるよう計算した．
- この手順で 10^5 個の処方を作成し，これらの処方を 4.2.2 で説明した最良モデルで予測した．

また，逆解析によって得られた処方の予測精度を検証するため，作成した処方セットからサンプリングを行い，実際の洗浄率を測定した．

表 4.4: 逆解析による処方設計の条件 (wt%)

Category	Material Name	Randomize the Addition Rate	Minimum Weight	Maximum Weight
Anionic surfactant	Potassium Cocoyl Glutamate	Yes	3	20
Amphoteric surfactant	Lauramidopropyl Hydroxysultaine	Yes	3	20
Nonionic surfactant	Eicosaglycerol Hexacaprylate; PEG-20 Glyceryl Triisostearate, etc.	Yes	0	10
Polyols	PPG-9 Diglycerol Ether; Glycerin, etc.	Yes	0	30
pH adjuster	Citric Acid	No	0.8% (fixed ratio)	
Base	Water	No	100- Σ (other material amount)	

4.3 結果と考察

4.3.1 AI のモデリング

構築した洗浄率を予測する各モデルの予測精度を表 4.5 に示す．最高の予測精度は $R^2 = 0.770$ で得られ，これは RMSE に換算して約 15% に相当する．本精度は実験の代替にはならないものの，スクリーニング用途（低性能の見込みが高い配合の実験を省くなど）に足る精度である．また，分子記述子の使用により，全てのモデルで予測精度が向上し，分子記述子として表現される化学的性質が洗浄率の予測を成功させたことを示している．最良モデルである Extra-Trees と分子記述子，ハンセン溶解度パラメータ，特徴抽出を使用した予測-実測散布図を図 4.3 に示す．

表 4.5: 各モデルの予測精度

	mol/weight Fraction	Descriptors	Hansen Solubility Parameters	Feature Extraction	Extra-Trees	Random Forest Regressor	SVR	Lasso	PLS
1	weight	Not used	Not used	Not used	0.661	0.652	0.488	0.526	0.503
2	mol	Not used	Not used	Not used	0.668	0.643	0.469	0.503	0.472
3	weight	Used	Not used	Not used	0.751	0.715	0.567	0.565	0.541
4	mol	Used	Not used	Not used	0.753	0.751	0.509	0.526	0.519
5	weight	Not used	Used	Not used	0.725	0.657	0.320	0.582	0.535
6	mol	Not used	Used	Not used	0.719	0.654	0.305	0.556	0.505
7	weight	Used	Used	Not used	0.748	0.711	0.568	0.582	0.543
8	mol	Used	Used	Not used	0.751	0.749	0.524	0.537	0.526
9	weight	Used	Used	Used	0.770	0.738	0.671	0.564	0.532
10	mol	Used	Used	Used	0.768	0.753	0.620	0.523	0.504

4.3.2 逆解析による処方設計と特徴量効果の可視化

逆解析による処方設計の洗浄率を作成し，原料種ごとに層別したものを図 4.4 と図 4.5 に示す．これらの図の明灰色と暗灰色のボックスはデータの四分位範囲を示し，その境界線は中央値を示す．また，上下の水平バーは隣接するボックスの幅の 1.5 倍以内のすべての値を示す．

図 4.4 では，Eicosaglycerol Hexacaprylate が中央値（各ボックスの中央水平線）と最高値（上部水平線）の両方の観点から最高の洗浄率を示すことが分かった．図 4.5 では，予測データを「ノニオン界面活性剤を使用しない処方」と「ノニオン界面活性剤を使用する処方」の 2 つに階層化し，さらにノニオン界面活性剤とポリオールとの相互作用を推定するために，ポリオール種でさらに層別した．

箱ひげ図でノニオン界面活性剤の有無による層別結果を比較すると，ノニオン界面活性剤の追加により洗浄率が向上することが類推される．また，ポリオール種による層別結果を

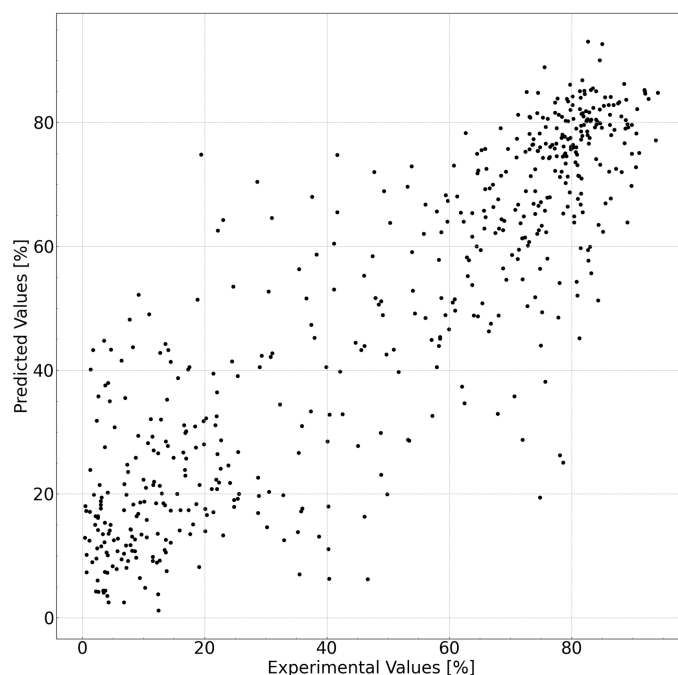


図 4.3: クレンジング剤処方の洗浄率の実験値 vs 予測値 (10-fold 交差検証各 fold の Test データ)

比較すると、疎水性の高いポリオール（例：PPG-9 Diglyceryl Ether, Cyclohexylglycerin）は、Glycerin よりも洗浄率をさらに高めることが示唆される。

逆解析による処方設計で得られた予測データを検証するため、逆解析で得られた処方からいくつかの処方を選択し、それらの洗浄率を測定した。処方と結果を表 4.6 および図 4.6 に示す。Eicosaglycerol Hexacaprylate をノニオン界面活性剤として、Cyclohexylglycerin もしくは PPG-9 Diglyceryl Ether をポリオールとして選択した処方（表 4.6 の Inv.A, Inv.B）は、実際の処方で高い洗浄率を示した。他のノニオン界面活性剤と Cyclohexylglycerin もしくは PPG-9 Diglyceryl Ether（表 4.6 の Inv.C, Inv.D）を含む処方はより低い洗浄率を示した。Glycerin を含む処方（表 4.6 の Inv.E, Inv.F）は、ノニオン界面活性剤の種類に関係なく低い洗浄率を示した。これらの傾向は図 4.4 と図 4.5 の結果に対応していた。予測精度は 70%以上の洗浄率と 15%未満の洗浄率の範囲でより良好であり、これは高洗浄率範囲と低洗浄率範囲の予測がより正確であった図 4.3 の結果にも対応していた。

表 4.6: 予測値と実際の洗浄率の比較用処方（逆解析例）

Category	Material Name	Inv.A	Inv.B	Inv.C	Inv.D	Inv.E	Inv.F
Anionic surfactant	Potassium Cocoyl Glutamate	6	7	13	8	8	7
Amphoteric surfactant	Lauramidopropyl Hydroxysultaine	5	9	4	9	7	7
Nonionic surfactants	Eicosaglycerol Hexacaprylate	10	10	0	0	10	0
	Decaglycerol Isostearate	0	0	10	0	0	0
	PEG-20 Glyceryl Triisostearate	0	0	0	10	0	10
Polyols	PPG-9 Diglyceryl Ether	0	11	0	11	0	0
	Cyclohexylglycerin	9	0	14	0	0	0
	Glycerin	0	0	0	0	11	13
pH adjuster	Citric Acid	0.8	0.8	0.8	0.8	0.8	0.8
Base	Water	69.2	62.2	58.2	61.2	63.2	62.2
Cleansing capability test							
Prediction / %		81.2	78.3	71.9	42.1	7.4	14.4
Actual / %		85.8	85.9	39.8	46.3	8.3	3.6

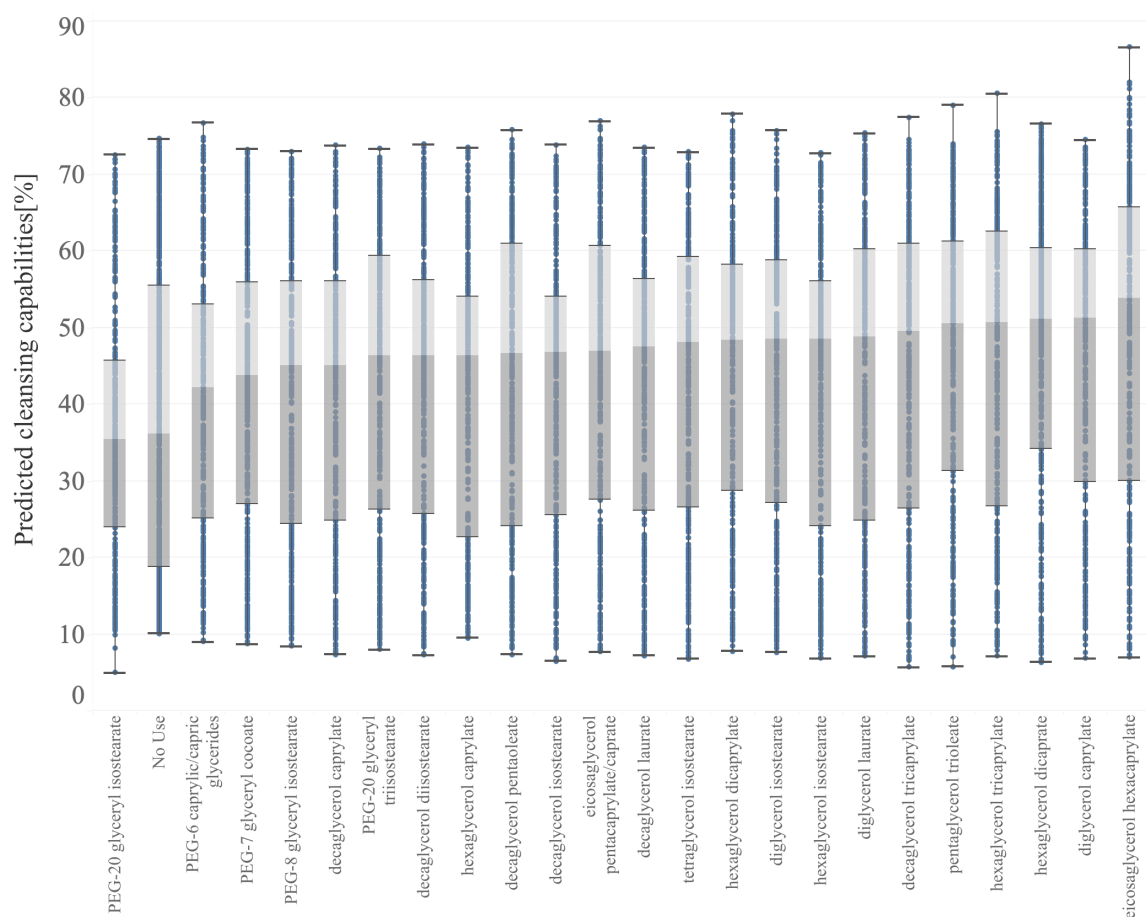


図 4.4: 逆解析による処方設計による洗浄率の箱ひげ図（ノニオン界面活性剤による層別）

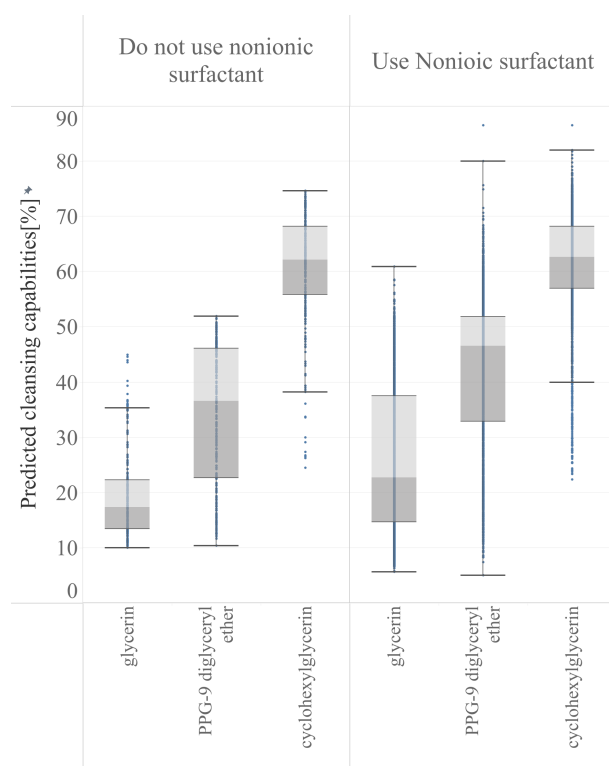


図 4.5: 逆解析による処方設計による洗浄率の箱ひげ図（ノニオン界面活性剤とポリオールによる層別）

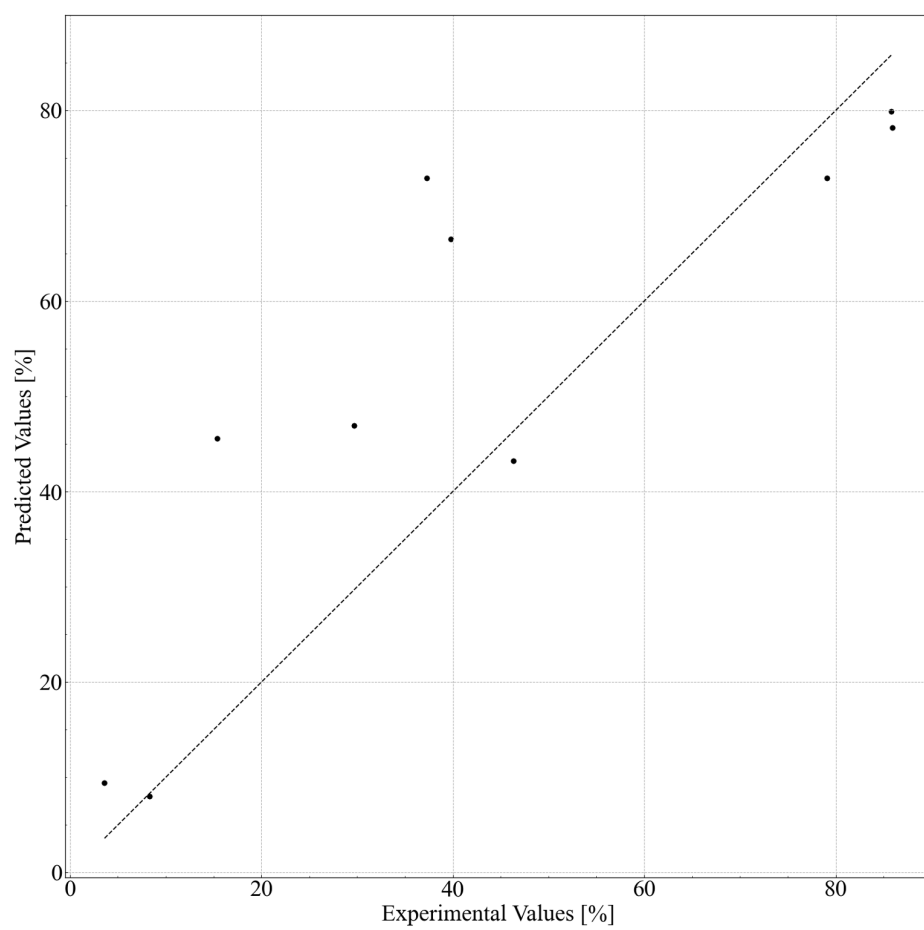


図 4.6: 逆解析処方における洗浄率の実験値と予測値の比較

4.4 考察

記述子の使用により、全てのモデルで予測精度が向上し、分子記述子として表現される化学的性質が洗浄率予測精度を向上させたことを示している。HSPs はいくつかのモデルで予測精度を改善したが、記述子を用いたモデルと比べると十分ではなかった。さらに、PEG-20 Glyceryl Triisostearate を用いた配合では性能が相対的に低く、性能の主要因はポリマー長そのものよりも、配合間相互作用に起因する高次構造（自己集合状態など）の寄与が大きい可能性が示唆された。

さらに、加重平均の計算において、線形ベースモデルでは weight% が適しており、決定木ベースモデルでは weight% と mol% の差は小さかった。mol% は化学量論の観点からより自然だが、本研究の処方では高分子量成分の影響で水が平均で 97mol% 以上を占めるため、mol% では水の影響が支配的になりやすく、線形ベースモデルはその影響を受けやすいと考えられる。

洗浄率の予測では、界面活性剤と水分子間の相互作用およびそれらの自己集合のため、非線形挙動も考慮する必要がある。決定木モデルは通常、非線形予測により適しているため、それらの予測精度は線形ベースモデルよりも高い。最良の方法は Extra-Trees であり、決定木ベースであるため、学習データに対して外挿領域となる推論では精度が低下しうる。その場合は追加実験によりデータを拡充し、再学習する必要がある。

逆解析による処方設計により、各材料が洗浄率に与える効果だけでなく、分子相互作用の結果としての材料の組み合わせによる効果も類推することができた。逆解析による処方設計はまた、より高い洗浄率を持つ処方の作成も支援した。

本研究では、予測方法と逆解析による処方設計方法を適用することにより、Eicosaglycerol Hexacaprylate と Cyclohexylglycerin/PPG-9 Diglycerol Ether で構成され、化粧汚れとしてのペンシルアイライナーの除去において 85% を超える高い洗浄率を示すクレンジング剤処方を特定した。

4.5 結論

機械学習を用いた人工知能（AI）により、界面活性剤の自己集合体と原料の化学的性質の効果を組み込んだ洗浄率予測システムを構築した。クレンジング性能の予測において $R^2 = 0.770$ (RMSE=15%) の精度を得た。この精度は、低洗浄率が予想される配合の事前除外などスクリーニング用途として十分である。処方中の化粧品原料間の相互作用である非線形挙動は、処方設計者がその性能を予測することをより困難にする。しかし、記述子による化学的性質を組み込むことにより高い精度が得られた。

原料の分子構造と界面活性剤の自己集合体に基づくこの AI 予測モデルは、重回帰などの従来のアプローチよりも高い精度と優れた性能を示した。逆解析による処方設計の使用により、処方設計者は最高の洗浄率を得るためにどの原料を選択すべきかの情報を得ることができる。予測モデルと逆解析による処方設計は、化粧品開発に要する労力の大幅な削減に貢献する可能性がある。

本研究で開発した AI 技術を活用したクレンジング剤性能予測システムは、従来の経験的アプローチの限界を克服し、科学的根拠に基づく効率的な処方設計を可能にした。特に、分子記述子による化学的性質の数値化と機械学習モデルの組み合わせにより、成分間の複雑な相互作用や非線形効果を考慮した高精度な性能予測が実現された。

また、逆解析による処方設計の導入により、実際の実験を行うことなく大量の処方候補を評価することが可能となり、開発期間の短縮とコスト削減を達成した。これらの成果は、化粧品業界における製品開発プロセスの革新と、消費者ニーズへの迅速な対応を可能にする重要な基盤技術として位置づけられる。

今後の展開として、本手法を他の化粧品カテゴリーや機能性材料の開発に適用することで、より広範囲な製品開発の効率化が期待される。また、AIモデルの継続的な改善と新たな分子記述子の導入により、予測精度のさらなる向上も見込まれる。

Chapter 5

クレンジング剤開発への AI と分子シミュレーション手法の活用

5.1 本章の目的

本章では、前章で述べた手法を拡張し、より製品開発ニーズの高い、原料の新規の組み合わせや新規原料利用にフォーカスした、AI と分子シミュレーションを統合した洗顔フォーム性能予測システムの構築について述べる。QSPR メソッドでモデル化が困難な多成分系である界面活性剤混合物の挙動予測における課題を克服するため、人工知能 (AI) と散逸粒子動力学 (DPD) の両方の力を活用した新しいアプローチを提案する。分子シミュレーションデータと機械学習を組み合わせることで、界面活性剤の自己組織化と成分の化学特性の影響を考慮したクレンジング剤による洗浄率予測システムを構築し、製薬、化粧品、材料科学など様々な分野で幅広い応用を持つ界面活性剤混合物システムに対する堅牢で正確な予測モデルの開発を目指す。この文脈での AI と分子シミュレーションの成功的統合は、多成分系の合理的設計と最適化における新しいパラダイムへの道を開く可能性がある。

5.2 実験手法

5.2.1 洗浄率の評価

クレンジング剤はイオン性界面活性剤、両性界面活性剤、ノニオン界面活性剤、ポリオール、pH 調整剤、水で構成される（使用原料例を表 5.1 に示す）。これらを、加熱しながら十分に混合・攪拌することにより調製した。

本章の目的変数となる洗浄率の測定手順の概略図を図 5.1 に示す。まず白色の人工皮革に防水性ペンシルアイライナーを塗布し、30 分間乾燥させた。その後、対応する洗顔フォームサンプル 0.1 mL を乾燥したペンシルアイライナー上に添加し、30 回擦った後、すすいで乾燥させた。洗浄率は、以下の式により色差から算出したペンシルアイライナーの残留率を色差にて評価した [15]：

洗浄率 =

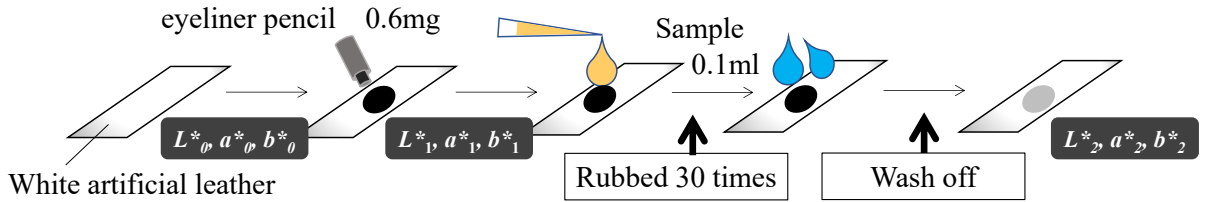
$$\frac{\sqrt{(L_0^* - L_1^*)^2 + (a_0^* - a_1^*)^2 + (b_0^* - b_1^*)^2} - \sqrt{(L_0^* - L_2^*)^2 + (a_0^* - a_2^*)^2 + (b_0^* - b_2^*)^2}}{\sqrt{(L_0^* - L_1^*)^2 + (a_0^* - a_1^*)^2 + (b_0^* - b_1^*)^2}} \quad (5.1)$$

表 5.1: 本研究で使用した原料カテゴリーの例

Category	The Number of Ingredients	Ingredient Examples
Anionic surfactant	8	Potassium Cocoyl Glutamate; Potassium Cocoyl Glycinate
Amphoteric surfactant	4	Lauramidopropyl Hydroxysultaine; Sodium Cocoamphoacetate
Nonionic surfactant	24	Eicosaglycerol Hexacaprylate; Decaglycerol Isostearate; PEG-20 Glyceryl Triisostearate
Polyol	33	PPG-9 Diglyceryl Ether; Cyclohexylglycerin; Glycerin
pH adjuster	1	Citric Acid
Base	1	Water

ここで, L^* は明度, a^* および b^* は色度を表し, $L^*a^*b^*$ は色彩計 (CM-2600d, Konica Minolta, Inc.) で測定した色空間である. (L_0^*, a_0^*, b_0^*) , (L_1^*, a_1^*, b_1^*) , (L_2^*, a_2^*, b_2^*) は, それぞれペンシルアイライナー塗布前, 塗布後, クレンジング後の白色人工皮革の色空間値を表す.

なお, 本章の洗浄率においては, 深層学習における活性化関数に sigmoid 関数も検討したため, 前章と異なり百分率にしていない. また, サンプル数は 430 である (シミュレーション数の制限より, 前章の 537 サンプルより少ない).



L^*, a^*, b^* value were measured by a colorimeter (CM-2600d, Konica Minolta, Inc.)

図 5.1: 本研究で調製した処方の洗浄率を決定する評価テストの概要

5.2.2 粗視化分子シミュレーション

我々は DPD シミュレーション手法 [57–59] を用いて, 多成分洗浄液の形態を再現する. DPD 法の基本方程式は Newton の運動方程式であり, 全 DPD 粒子に作用する 3 種類の力 (保存力 ($\mathbf{F}^C$), 散逸力 ($\mathbf{F}^D$), ランダム力 ($\mathbf{F}^R$)) で構成される. 粒子 i に対する Newton の運動方程式は次の通りである:

$$m_i \frac{d\mathbf{v}_i}{dt} = \mathbf{f}_i = \sum_{j \neq i} \mathbf{F}_{ij}^C + \sum_{j \neq i} \mathbf{F}_{ij}^D + \sum_{j \neq i} \mathbf{F}_{ij}^R \quad (5.2)$$

ここで, m は粒子の質量, $\mathbf{v}$ は粒子の速度を表す. 一方, 保存力 $\mathbf{F}^C$ は以下の式を用いて計算される.

$$\mathbf{F}_{ij}^C = \begin{cases} a_{ij} \left(1 - \frac{|\mathbf{r}_{ij}|}{r_c}\right) \mathbf{n}_{ij}, & |\mathbf{r}_{ij}| \leq r_c \\ 0, & |\mathbf{r}_{ij}| > r_c \end{cases}, \quad (5.3)$$

ここで, $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$, $\mathbf{n}_{ij} = \mathbf{r}_{ij}/|\mathbf{r}_{ij}|$, r_c は力の有効範囲のカットオフ距離を指定する. 相互作用パラメータ a_{ij} は粒子 i と j 間の反発力の大きさを決定し, $a_{ij} = a_{ii} + 3.268 \chi_{ij}$ で表され, $a_{ii} = 62.38 k_B T / r_c$ [60] であり, Flory-Huggins パラメータ χ を含む. 本研究で使った Mesoscopic モデルの Flory-Huggins パラメータを決定するため, 以下の式を用いた: [61, 62]

$$\chi_{ij} = \frac{V_{ij}}{k_B T} (\delta_i - \delta_j)^2 \quad (5.4)$$

ここで, V_{ij} はビーズ i と j の平均モル体積, k_B は Boltzmann 定数, δ_i はビーズ i の溶解度パラメータである. 溶解度パラメータ δ およびモル体積 V は, J-OCTA シミュレーションソフトウェア [63] を用いた分子動力学 (MD) シミュレーションにより各ビーズの凝集エネルギーを計算することで得られた.

MD 計算には完全原子モデルを使用し, 力場値には一般 Amber 力場 (GAFF [64]) を, 電荷値には分子軌道計算を用いた. 隣接粒子間の結合は調和ばねで接続され, その力場は以下のように定義される: $\mathbf{F}_{ij}^S = -k_s (1 - |\mathbf{r}_{ij}|/r_s) \mathbf{n}_{ij}$

ここで, $r_s = 0.86 r_c$ は平衡結合距離, $k_s = 100 k_B T$ はばね定数を表す. シミュレーションにおける粒子密度は $\rho = 5 r_c^{-3}$ と設定され, シミュレーションボックスの体積は $30 \times 30 \times 30 r_c^3$ で, 3次元すべてに周期境界条件を適用した. 430 種類の洗浄液組成に対する DPD シミュレーションを実施し, その結果, 各原料組成に応じて, ミセル, ベシクル, 層状構造など様々な平衡形態が得られた.

5.2.3 粗視化モデルとシミュレーションからの特徴抽出

機械学習アプリケーション向け特徴抽出プロセスの一環として, 粗視化分子モデルと DPD シミュレーション結果の両方から情報を収集した. 洗浄の鍵となる因子である両親媒性に関して分子の特性を決定するため, 溶解度パラメータ δ を用いた. 両性分子の自己組織化挙動は, 親水性基と疎水性基の比率など, その両性度によっても強く影響を受ける [65–67]. したがって, 各分子に対して, 親水性ビーズの割合 (f_{HI}) と両親媒性度合い ($\Delta\delta$) の2つのパラメータを定義した. f_{HI} は, 粗視化モデルにおける親水性 (HI) ビーズの割合として計算される. 一方, $\Delta\delta$ は疎水性 (HO) ビーズと親水性 (HI) ビーズの平均溶解度パラメータの差 $\Delta\delta = \delta_{\text{HI}} - \delta_{\text{HO}}$ として定義される.

HI beads は溶解度パラメータ $\delta_{\text{HI}} \geq 24 \sqrt{k_B T / r_c^3}$ を満たすものと定義され, HO beads は $\delta_{\text{HO}} < 24 \sqrt{k_B T / r_c^3}$ で特徴づけられた. この閾値は, 溶解度パラメータが約 24 であるポリエチレングリコール (PEG) 単量体が親水性として分類されるように設定された. HI beads および HO beads の分類 (HI ビーズ定義) を図 5.2 に示す.

第二の特徴抽出として, DPD シミュレーションから得られた洗浄溶液の平衡形態を定量的に評価するいくつかのパラメータとして, 系のポテンシャルエネルギー U と圧力 P を導入した. 系の U は, 全ての相互作用に対する保存力の積分によって計算された. 圧力 P は, 圧力テンソル (ビリアル定理により計算) の対角成分を平均化して $P = (P_{\text{xx}} + P_{\text{yy}} + P_{\text{zz}})/3$ に従い計算した. これらの値を用い, 洗浄溶液の様々な自己組織化形態をコンピュータが読み取れる形式に定量的に符号化した.

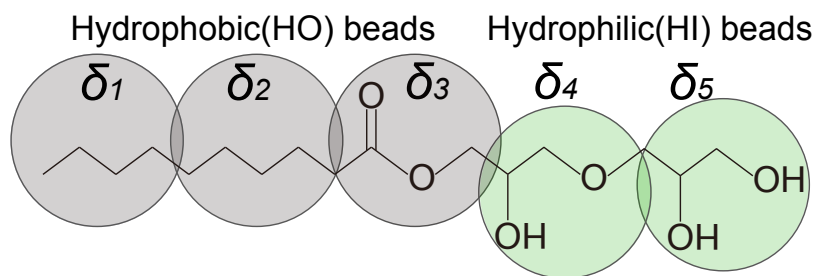


図 5.2: 粗視化分子モデルにおける溶解度パラメータ δ に基づく親水性 (HI) ビーズと疎水性 (HO) ビーズの分類概念図. 親水性ビーズの割合 (f_{HI}) と両親媒性の度合い ($\Delta\delta$) は各ビーズの溶解度パラメータ (δ) から算出する. 本例では $f_{\text{HI}} = 2/5$, $\Delta\delta = [(\delta_4 + \delta_5)/2 - (\delta_1 + \delta_2 + \delta_3)/3]$.

5.2.4 機械学習

5.2.4.1 データフロー

本章では、処方データそのものによる AI の学習に対して、化学的相互作用を組み込むため、前述のモデリングパラメータおよび DPD シミュレーションから得られた記述子を用いた AI のモデリングも組み込んだ。データ処理フローチャートを図 5.3 に示す。

本研究で作成した記述子の有用性を検証するため、学習データとして、(i) 原料の比率表として表現される配合データ、(ii) モデリングパラメータから導出された記述子を伴う配合データ、(iii) シミュレーションから導出された記述子で補完された配合データ、(iv) モデリングとシミュレーションの両方からの記述子を統合した配合データの 4 つのデータパターンを定義した。

5.2.4.2 モデリング

配合-特性の関係を捉えるため、多様な機械学習モデル群を採用した。まず、階層的決定構造を通じて非線形相互作用を捕捉するのに優れた、Random Forest Regressor や Extremely Randomized Trees (Extra-Trees, ET) Regressor などのツリーベースのアンサンブル手法を採用した。また、線形モデルでは、疎な特徴量選択のための Lasso 回帰と次元削減のための Partial Least Squares (PLS) を採用した。加えて、複雑な非線形境界に対応するため、Radial 基底関数カーネルを用いた Support Vector Regressor (SVR) も採用した。

深層学習として、全入力特徴量を統一されたネットワーク構造で処理する Fully Connected Neural Network アーキテクチャを実装した。モデルは入力層に続き、過学習防止のため ReLU 活性化関数とドロップアウト正則化を備えた複数の隠れ層で構成される。また、各隠れ層は、処方データ、分子モデリング記述子、シミュレーション由来記述子を含む全入力ソースからの統合特徴量を処理する。処理された特徴量は追加の隠れ層群を経て、線形活性化関数による単一のスカラ出力へと集約される。モデルは平均二乗根誤差損失関数のもと、Adam Optimizer で最適化した。

深層学習に加え、処方表現におけるシーケンス様依存性を捕捉するため、Gated Recurrent Unit (GRU) アーキテクチャも採用した。各処方は固定長シーケンスとして符号化され、位置は非ゼロの原料スロットに対応し、最大シーケンス長に合わせるためパディングが適用された。各位置では、複数の特徴チャンネルが原料使用量、分子モデル由来の記述子、シミュレーション由来の記述子を統合した。シーケンス処理には一方向 GRU を採用し、最終隠れ状態は線形層を通じてスカラ予測値である洗浄率にマッピングされた。モデルはシャッフルされたミニバッチを用い、平均二乗根誤差を最小化する Adam Optimizer で学習された。

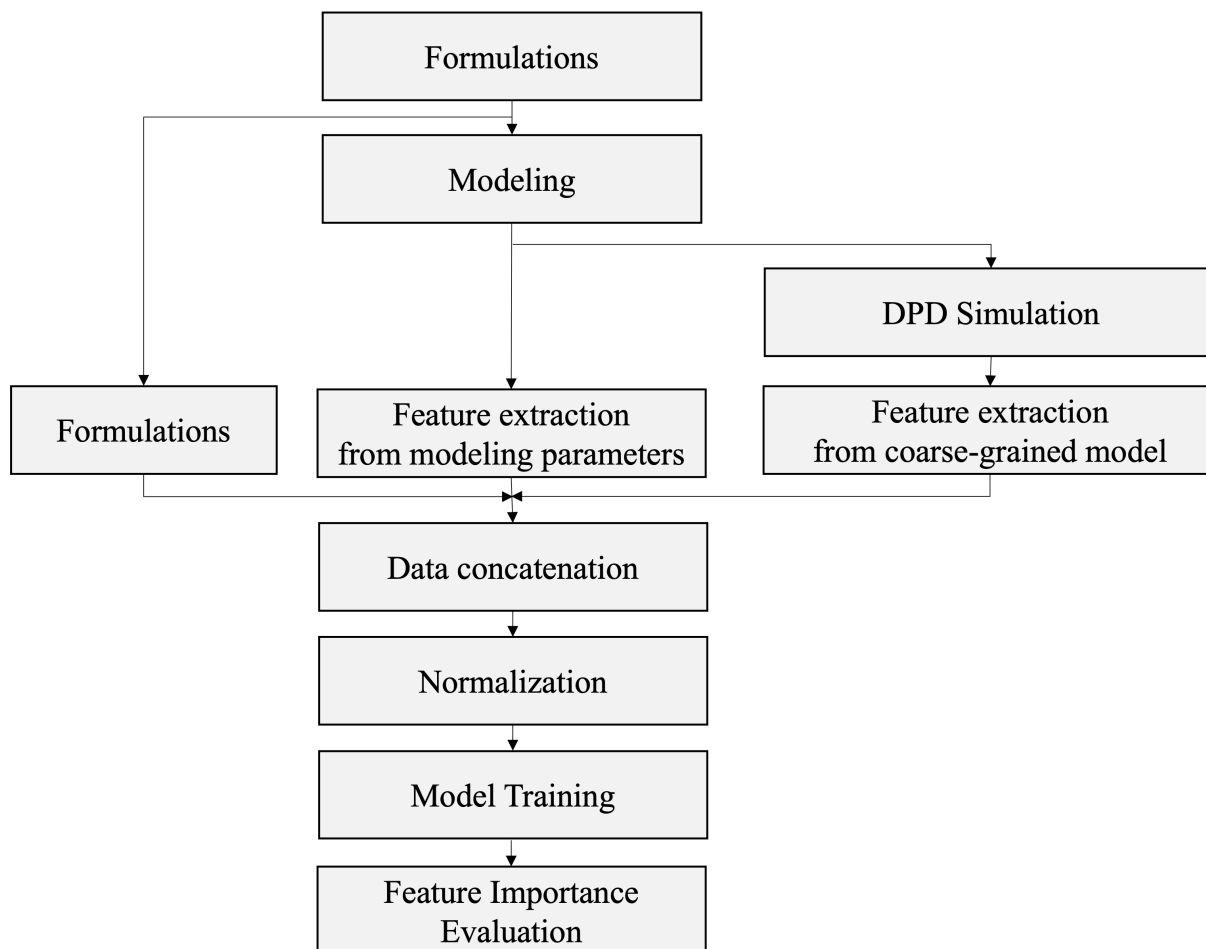


図 5.3: データ処理フロー

全ての説明変数は平均 0, 標準偏差 1 で正規化されたが, テストデータのリーケージを防ぐため, 各交差検証 fold ごとに独立して正規化を実施した. 正規化パラメータ (平均と標準偏差) は学習データのみから計算され, 各 fold 内の学習データとテストデータの両方に適用された. 全モデルのハイパーパラメータ範囲を表 5.2 に示した. 深層学習モデルの実装には TensorFlow フレームワークを, 再帰型ニューラルネットワークの実装には PyTorch フレームワークを使用した. その他の機械学習には Scikit-Learn ライブラリを使用した.

5.2.4.3 モデル評価

交差検証におけるデータのグループ化を体系的に制御するため, 異なる分割戦略を適用した:

- **Points Out** (ランダム分割): 交差検証フレームワーク内で全データを各 fold にランダムに割り当てた. この手法は標準的な n 分割交差検証に相当し, データを評価用に n グループにランダム分割する.
- **Mixtures Out** (混合物組成除外法): 同一構成材料 (同一の原料集合) を異なる比率で配合した混合物をグループ化し, 同一の外部 fold に割り当てた. これにより, 特定の混合物配合から得られた全サンプルが, 学習データセットまたはテストデータセットのいずれか一方にのみ属し, 両方に属さないことが保証される.

表 5.2: モデリングにおけるハイパーパラメータ設定

Model Name	Hyperparameter	Value
Extremely Randomized Tree (Extra-Trees) regressor	n_estimators	200, 500
	max_depth	None, 15, 30
	max_features	0.1, 0.3, 0.5, 1.0
	min_samples_split	2, 5
	min_samples_leaf	1, 2, 4
	bootstrap	True, False
RandomForestRegressor	n_estimators	200, 500
	max_depth	None, 15, 30
	max_features	0.1, 0.3, 0.5, 1.0
	min_samples_split	2, 5
	min_samples_leaf	1, 2, 4
	bootstrap	True, False
SVR	C	0.001, 0.01, 0.1, 1, 10, 100, 1000
	kernel	rbf
	gamma	0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10
PLS	n_components	1-10
Lasso	alpha	0.001, 0.01, 0.1, 1
Neural Network	learning_rate	0.001, 0.01
	dropout_rate	0.3
	hidden_layers	2
	hidden_units	100%, 50% of input size
	batch_size	32
	epochs	100, 150
	activation	relu
RNN (GRU)	hidden_size	64, 128
	num_layers	2
	batch_size	32
	epochs	100, 150
	learning_rate	0.001, 0.01
	dropout_rate	0.3

Note: Machine learning models (Extra-Trees, Random Forest Regressor, Support Vector Regressor, Partial Least Squares, Lasso), deep learning models (Neural Network), and recurrent neural networks (RNN) hyperparameter values are shown.

- **Compounds Out** (成分除外) : 同一ポリオールを含む混合物をグループ化し、同一外部 fold に割り当てた。なお、1 処方に 2 つのポリオールが含まれる処方については、これも独立したグループとして割り当てた。これにより、いずれのテスト fold にも、対応する学習セットに存在しないポリオールを含む混合物配合が少なくとも 1 つ含まれることが保証された。なお、アニオン性、両性およびノニオン性界面活性剤については、使用される原料の種類に偏りがあるため、Compounds Out は採用しなかった。

全モデルの検証は、各評価戦略に対してネストされた交差検証戦略を用いて実施された。10 分割の外側交差検証 (Outer Cross Validation) ループを用いて、データセットを学習セットとテストセットに分割した。モデルの性能は、外側 CV 学習データと外側 CV テストデータの 2 つのデータセットに対して個別に評価された。このアプローチは、モデルの汎化能力を包括的に評価するために行われた。

各モデルの学習フェーズでは、主要評価指標として平均二乗誤差 (MSE) を用いてハイパーパラメータ調整を実施した。学習データ内で 9 分割の内側交差検証 (Inner CV) ループを実行し、グリッドサーチによるハイパーパラメータ最適化を実施し、検証損失に基づき最適なハイパーパラメータ設定を選択した。最適ハイパーパラメータ決定後、モデルは完全な学習セットを用いて再学習され、Outer CV のテストセットで評価された。なお、分割乱数を変えても結果が変わらないことを別途確認している。

学習時の最適化損失関数には MSE を用いたが、最終モデルの性能提示には解釈性を高めるため R^2 と RMSE を用いた。各指標の定義は以下の通り：

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5.5)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5.6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5.7)$$

ここで、 $\hat{y}_i$, y_i , および $\bar{y}$ は、それぞれ予測値、実測値、および実測出力の平均値である。

5.2.4.4 モデリング、シミュレーション記述子追加効果の統計的分析

予測精度に対する追加記述子 (分子モデリングおよびシミュレーション記述子) の組み込み効果を定量的に評価するため、包括的な統計分析を実施した。本分析では、2 組のデータ間の差異を比較しその有意性を判定する統計手法である対応のある差の検定を用いた。記述子追加効果に対するロバストネスな評価を確保するため、この解析は各交差検証戦略ごとに独立して実施された。

各交差検証 fold について、ベースラインモデル (原料データのみモデル) と追加記述子で強化されたモデル間の RMSE の差を計算した：

$$d_m = \text{RMSE}_{\text{baseline}} - \text{RMSE}_{\text{with_modeling_descs}} \quad (5.8)$$

$$d_s = \text{RMSE}_{\text{baseline}} - \text{RMSE}_{\text{with_simulation_descs}} \quad (5.9)$$

$$d_{m+s} = \text{RMSE}_{\text{baseline}} - \text{RMSE}_{\text{with_modeling_and_simulation_descs}} \quad (5.10)$$

正の値は性能の向上を示す。このペア分析は、全特徴条件で一貫した乱数シードを維持し、各 fold が同一のデータ分割を含むことを保証することで実現した。

また、学習モデルの差異を吸収するため、最小二乗法（OLS）回帰もあわせて採用した。モデル式は以下のとおりである：

$$\begin{aligned} d_{ijk} = & \beta_m \cdot I(\text{condition}_i = \text{modeling}) + \beta_s \cdot I(\text{condition}_i = \text{simulation}) \\ & + \beta_{m+s} \cdot I(\text{condition}_i = \text{modeling and simulation}) \\ & + \sum_{j=1}^J \gamma_j \cdot I(\text{model}_j) + \epsilon_{ijk} \end{aligned}$$

ここで d_{ijk} は特徴条件 i ，モデル j ，fold k における差を表す。モデルには各機械学習アルゴリズム固有の特性を制御するための固定効果を含みつつ、特徴強化効果に焦点を当てるため fold 効果は除外した。

各特徴条件について、帰無仮説 $H_0 : \beta_i = 0$ に対し対立仮説 $H_1 : \beta_i > 0$ を検定した。3つの特徴条件（モデリングデータ、シミュレーション記述子、モデリングとシミュレーション記述子）にわたる多重比較問題に対処するため、ファミリー全体の誤り率を $\alpha = 0.05$ で制御するために Holm のステップダウン手順を適用した。残差における潜在的な不均一分散を考慮するため、標準誤差は HC3 ロバスト共分散推定を用いて推定した。

この統計的枠組みは、3つの交差検証戦略である Points Out（全体的なモデル性能評価のための標準交差検証）、Mixtures Out（未知の混合物原料への汎化性能評価）、Compounds Out（未観測原料（本研究ではポリオール）への汎化性能評価）に独立して適用された。各戦略の独立した分析により、異なる汎化シナリオにおける特徴量強化効果を包括的に評価でき、追加記述子が予測性能向上に寄与するタイミングと方法を明らかにする。

5.2.4.5 変数の影響分析

本研究では、洗浄率に最も大きく寄与する変数を予測するため、SHAP（SHapley Additive exPlanations）を採用した。SHAP はゲーム理論に基づく高度なツールであり、各特徴量に重要度値を割り当てることで機械学習モデルの出力結果を解明する [49]。この値は各特徴量がモデルの予測に与える影響度を表し、複雑なモデルの解釈を容易にする。

本分析では、洗浄率予測への影響を特定するため、原料添加率を除外し、主にモデリングおよびシミュレーションパラメータに焦点を当てた。SHAP の採用により、モデルの予測を分解することが可能となり、拡張された処方枠組みにおける各特徴量の寄与度を理解しやすくなった。

入力特徴量の構造をさらに解釈し、潜在的なパターンを明らかにするため、同じモデリングおよびシミュレーション由来の記述子セットに対して主成分分析（PCA）を実施した。主成分分析（PCA）は、相関のある変数を相関のない主成分の集合に変換する次元削減手法であり、データ内の支配的な傾向を可視化することを可能にする。本研究では、PCA を用いて特徴空間内の直交する変動軸を特定し、異なる物理化学的記述子が互いにどのようにクラスタリングまたは対立するかを決定した。

初期の2つの主成分に対する各特徴量の負荷量を詳細に検討することで、データの分散に同様または反対の寄与をする変数群を効果的に分類できた。この分析は、SHAP に基づく特徴量重要度に対して教師なしの補完的視点を提供し、特徴量間の内部関係と洗浄率予測におけるそれらの集合的挙動の理解を深めるのに役立った。

5.3 結果と考察

5.3.1 シミュレーションから導出された記述子

各製剤について、メゾスケール自己組織化挙動を捉えるため DPD シミュレーションを実施した。代表的なスナップショットを図 5.4 に示す。これらのスナップショットからポテンシャルと圧力を抽出し、後述する機械学習および深層学習モデルにおけるシミュレーション由来記述子として用いた。

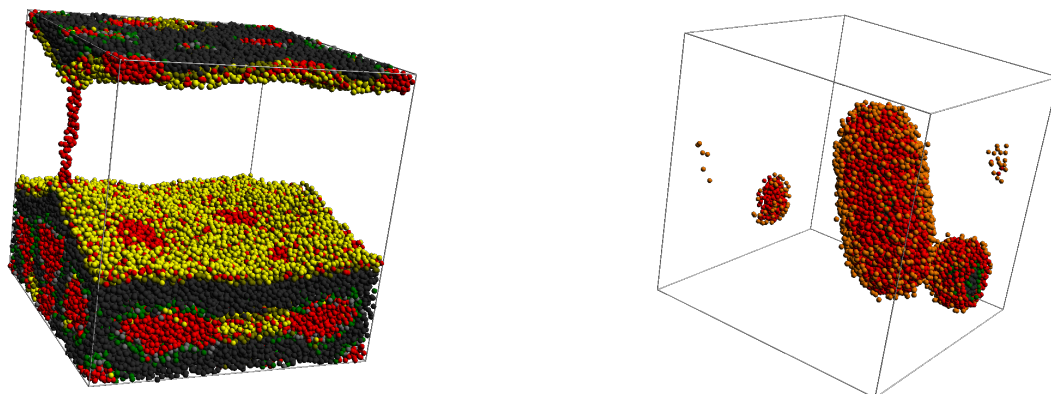


図 5.4: AI 入力用の特徴抽出に用いた代表的な散逸粒子動力学 (DPD) スナップショット。

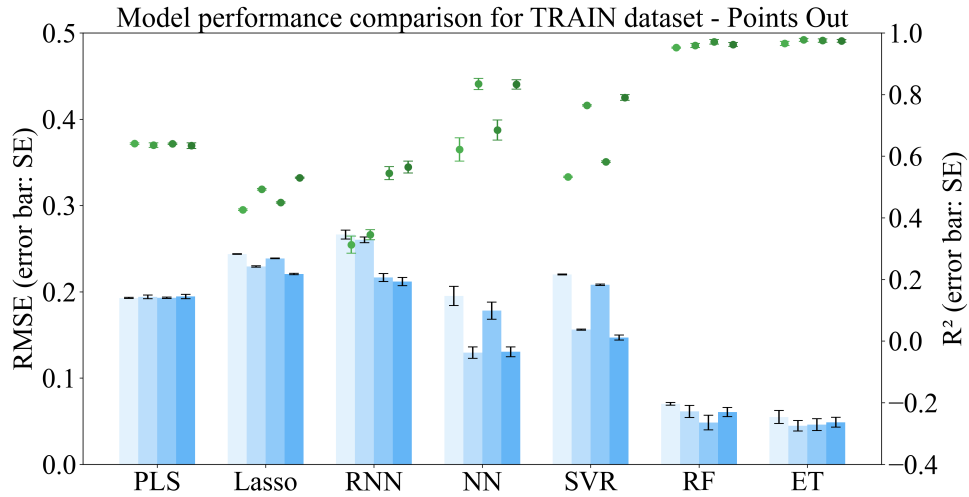
5.3.2 機械学習

洗浄率を予測する AI モデルを構築した。各モデルの予測精度は、図 5.5, 図 5.6, 図 5.7 に示す。Points Out, Mixtures Out, Compounds Out の各分野における最良モデルの予実プロットは、それぞれ図 5.8, 図 5.9, 図 5.10 に示されている。

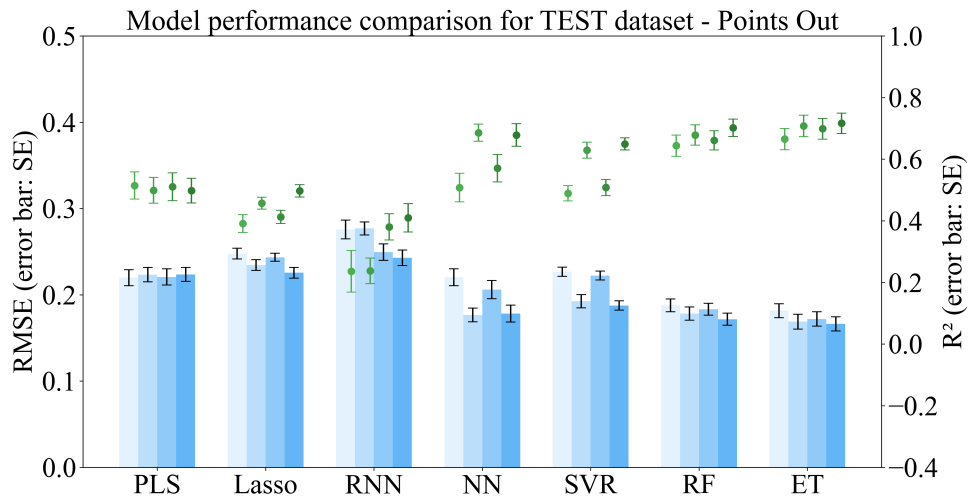
精度順序は一貫して「Points Out > Mixtures Out > Compounds Out」と観察され、新規組み合わせや新規原料特性の予測における課題を指摘している。重要なことに、すべての交差検証スキーム (Points Out, Mixtures Out, Compounds Out) において、モデリングおよび／またはシミュレーション記述子を組み込むことで精度が向上した。これは表 5.3 (a) および (b) にまとめられている。

Points Out 試験において、配合、モデリング、シミュレーション記述子を組み込んだ場合に最高精度が達成され、決定係数 R^2 は 0.716 であった。Mixtures Out 試験では、決定係数 R^2 が 0.573 で最高精度が達成された。Compounds Out では、決定木ベースの手法 (Random Forest Regressor モデルおよび Extra-Trees を含む) が高い精度を示した。様々なアルゴリズムの中で、Extra-Trees は Points Out および Mixtures Out の検証において最も優れた性能を示した一方、Random Forest Regressor は Compounds Out の検証において最高レベルの精度を示した。

混合物の場合、構成成分間の相互作用が非線形効果を引き起こし、洗浄率の予測を複雑化する。この複雑性は多成分混合物において増大する。実際、界面活性剤は水との相互作用を含む分子間相互作用や電荷効果の影響を受け、ミセルやベシクルなどの高次構造を形成する。したがって、配合物の添加比率のみに基づいて混合物を予測することは大きな課題となる。この文脈において、分子モデリングおよび DPD シミュレーション由来の記



(a) 学習データセットのモデル性能比較 (Points Out)



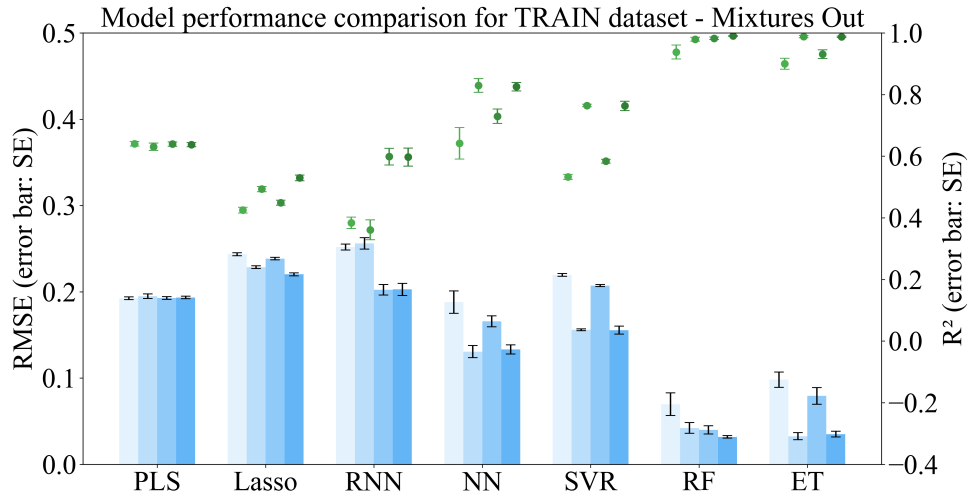
(b) テストデータセットのモデル性能比較 (Points Out)

- RMSE by formulation data
- RMSE by formulation and modeling data
- RMSE by formulation and simulation data
- RMSE by formulation, modeling, and simulation data
- R² by formulation data
- R² by formulation and modeling data
- R² by formulation and simulation data
- R² by formulation, modeling, and simulation data

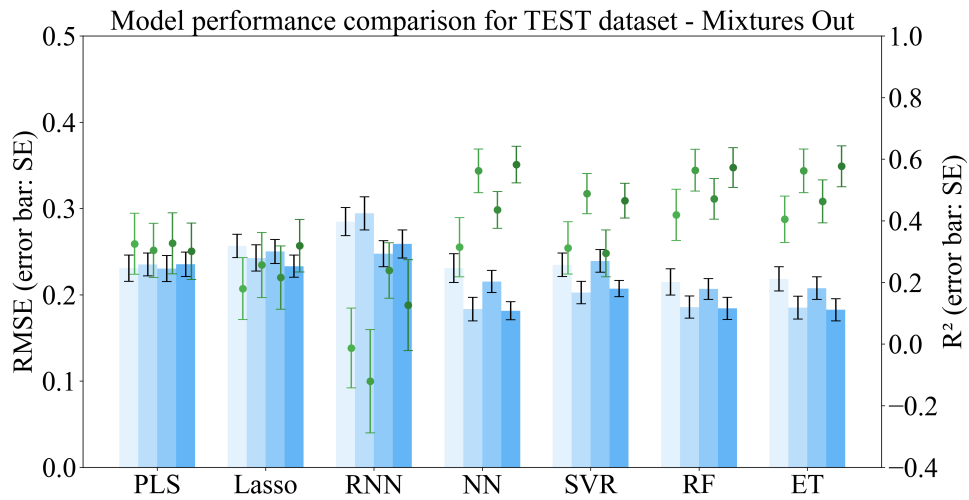
(c) モデル性能比較の凡例

図 5.5: 各モデルの性能比較 (Points Out 検証).

モデル略称: PLS = Partial Least Squares; RNN = Recurrent Neural Network (GRU); NN = Fully Connected Neural Network; SVR = Support Vector Regressor; RF = Random Forest Regressor; ET = Extra-Trees.



(a) 学習データセットのモデル性能比較 (Mixtures Out)



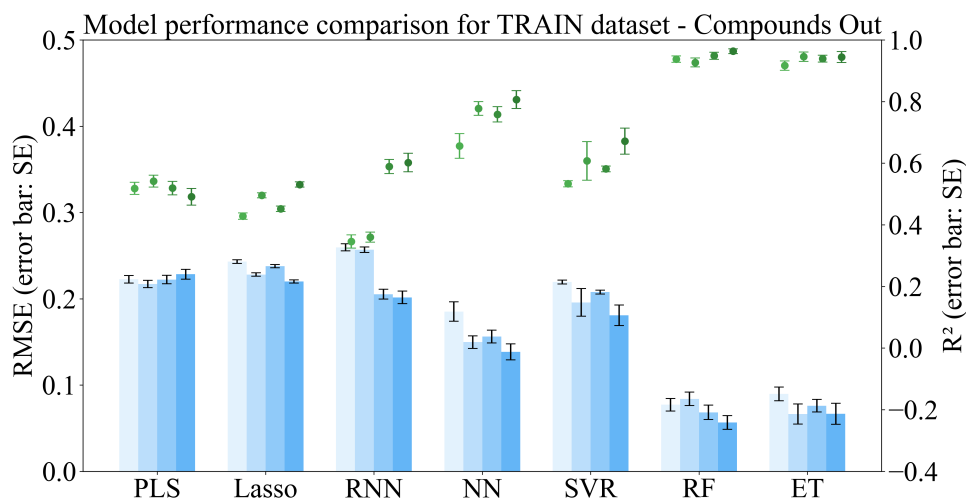
(b) テストデータセットのモデル性能比較 (Mixtures Out)

- RMSE by formulation data
- RMSE by formulation and modeling data
- RMSE by formulation and simulation data
- RMSE by formulation, modeling, and simulation data
- R² by formulation data
- R² by formulation and modeling data
- R² by formulation and simulation data
- R² by formulation, modeling, and simulation data

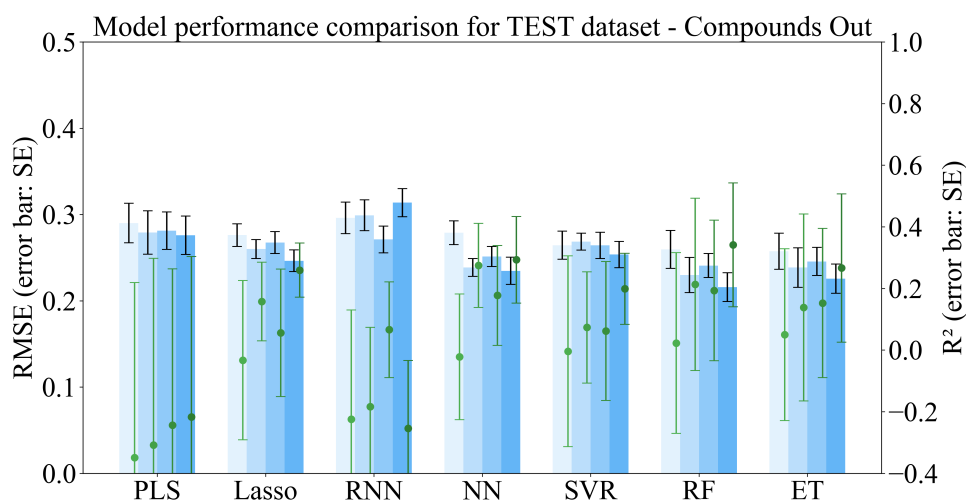
(c) モデル性能比較の凡例

図 5.6: 各モデルの性能比較 (Mixtures Out 検証).

モデル略称: PLS = Partial Least Squares; RNN = Recurrent Neural Network (GRU); NN = Fully Connected Neural Network; SVR = Support Vector Regressor; RF = Random Forest Regressor; ET = Extra-Trees.



(a) 学習データセットのモデル性能比較 (Compounds Out)



(b) テストデータセットのモデル性能比較 (Compounds Out)

- RMSE by formulation data
- RMSE by formulation and modeling data
- RMSE by formulation and simulation data
- RMSE by formulation, modeling, and simulation data
- R² by formulation data
- R² by formulation and modeling data
- R² by formulation and simulation data
- R² by formulation, modeling, and simulation data

(c) モデル性能比較の凡例

図 5.7: 各モデルの性能比較 (Compounds Out 検証).

モデル略称: PLS = Partial Least Squares; RNN = Recurrent Neural Network (GRU); NN = Fully Connected Neural Network; SVR = Support Vector Regressor; RF = Random Forest Regressor; ET = Extra-Trees.

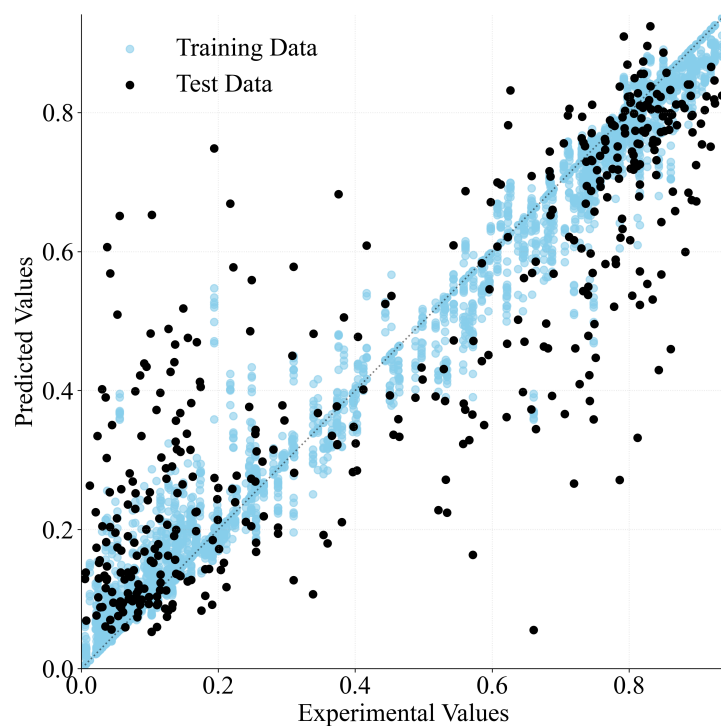


図 5.8: 予測値と実測値の相関散布図 (Points Out 検証の最良モデル : Extra-Trees, 配合+モデリング+シミュレーション記述子). 各点は 10 個の検証 fold におけるテスト予測群を表す.

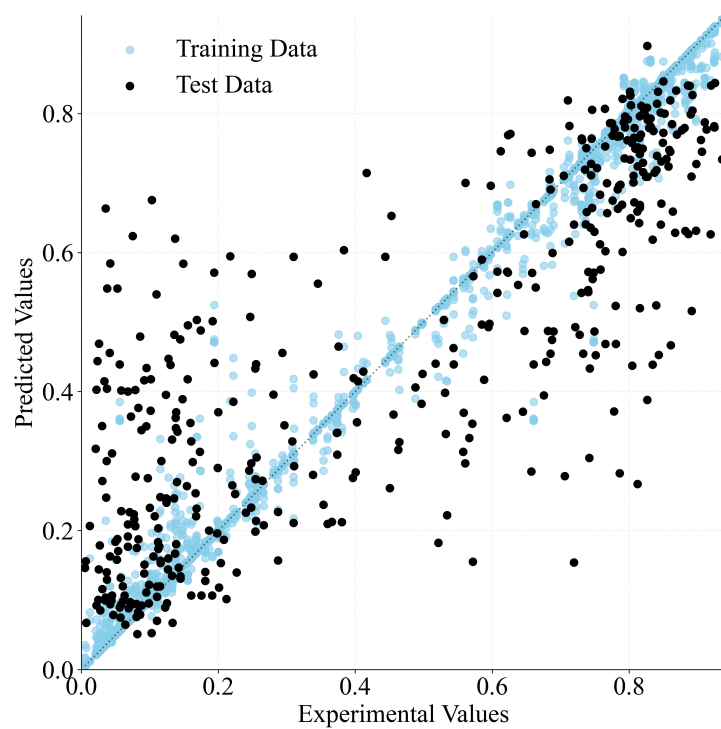


図 5.9: 予測値と実測値の相関散布図 (Mixtures Out 検証の最良モデル : Extra-Trees, 配合+モデリング+シミュレーション記述子). 各点は 10 個の検証 fold におけるテスト予測群を表す.

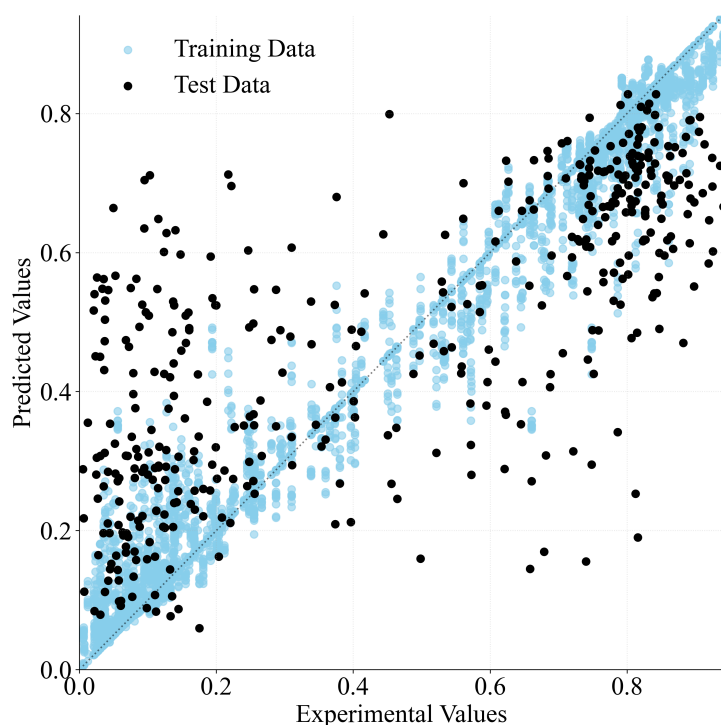


図 5.10: 予測値と実測値の相関散布図 (Compounds Out 検証の最良モデル: Random Forest Regressor, 配合+モデリング+シミュレーション記述子). 各点は10個の検証 fold におけるテスト予測群を表す.

表 5.3: モデル性能の要約: (a) R^2 , (b) RMSE による予測精度を各評価戦略で示す. 各 CV 法の最良モデルは, Points Out と Mixtures Out が Extra-Trees, Compounds Out が Random Forest Regressor.

(a) 異なるデータセットと交差検証法における R^2 の予測精度 (平均 $\pm$ 標準誤差).

	Points Out	Mixtures Out	Compounds Out
Formulation only	0.665 ± 0.034	0.420 ± 0.083	0.023 ± 0.293
Formulation + Modeling descriptors	0.708 ± 0.035	0.564 ± 0.068	0.213 ± 0.279
Formulation + Simulation descriptors	0.699 ± 0.034	0.471 ± 0.066	0.193 ± 0.228
Formulation + Modeling + Simulation descriptors	0.716 ± 0.033	0.573 ± 0.064	0.341 ± 0.201

(b) 異なるデータセットと交差検証法における RMSE の予測精度 (平均 $\pm$ 標準誤差).

	Points Out	Mixtures Out	Compounds Out
Formulation only	0.182 ± 0.008	0.215 ± 0.015	0.260 ± 0.022
Formulation + Modeling descriptors	0.169 ± 0.009	0.186 ± 0.013	0.230 ± 0.020
Formulation + Simulation descriptors	0.172 ± 0.008	0.207 ± 0.012	0.241 ± 0.014
Formulation + Modeling + Simulation descriptors	0.166 ± 0.008	0.184 ± 0.013	0.216 ± 0.017

述子の導入はいずれも有益であり、両者を併用した複合記述子で最大の精度向上が得られた (表 5.3 (a) および (b)).

精度向上の程度は、Points Out 法と比較して Mixtures Out 法でより顕著であった。この観察結果は、モデリングとシミュレーションから導出された記述子が、配合データだけでは捉えられない非線形情報を付与することを示唆している。Mixtures Out では R^2 が 0.420 から 0.573 に改善したのに対し、Points Out では 0.665 から 0.716 に増加した。これらの結果は、追加記述子が配合データのみでは捉えにくい情報を提供し、特にモデリングとシミュレーションを併用した複合記述子で最大効果が得られることを示唆している。Mixtures Out における単独効果を見ると、ベースライン (0.420) に対してモデリング記述子単独は 0.564 ($\Delta R^2 = +0.144$)、シミュレーション記述子単独は 0.471 ($\Delta R^2 = +0.051$)、複合記述子は 0.573 ($\Delta R^2 = +0.153$) であった (表 5.3 (a))。したがって、モデリングの寄与が相対的に大きい一方、シミュレーションも追加的な改善をもたらしており、両者が補完的に作用したと解釈できる。Compounds Out カテゴリーでは、全記述子の組み込み後、精度が初期値 0.023 から 0.341 へ向上した。この知見は予測能力向上の可能性を裏付ける。これらの精度は物理実験結果を完全に代替するには不十分であるが、実験的スクリーニング目的に叶う精度である。さらに、シミュレーション由来の記述子は分布外設定においても一定の改善を示した。処方データのみと比較し、シミュレーションデータのみ条件では、Mixtures Out 検証においてテスト R^2 が $\Delta R^2 = +0.051$ 増加し、Compounds Out 検証では $\Delta R^2 = +0.170$ 増加した。

決定木ベースモデル、特に Points Out および Mixtures Out 検証における Extra-Trees、ならびに Compounds Out 検証における Random Forest Regressor の優れた性能は、配合物内の非線形相互作用を捕捉する能力に起因すると考えられる。

また、洗浄率に対する複雑な分子間相互作用や構造的自己組織化現象の影響を考慮すると、Lasso や Partial Least Squares (PLS) などの線形モデルではこれらの効果を適切に捉えきれない。一方、決定木ベースの手法は特徴間の階層的関係を識別する高い能力を示し、このようなシステムのモデリングに特に有効である。

これらの知見は、非線形依存性が特に顕著な多成分系における複雑な物理化学的挙動の予測において、アンサンブル決定木ベースモデルを活用することの重要性を強調している。

5.3.2.1 モデリング、シミュレーション記述子追加効果の統計的分析

統計解析により、3つの交差検証戦略において追加記述子を組み込むことで予測精度が向上することが明らかになった。ただし、その効果は条件によって異なった。表 5.4 は、対比差分析による主要な知見をまとめたものである。

Points Out 戦略においては、分子モデリング記述子および複合記述子 (モデリングとシミュレーション) が有意な改善を示した (それぞれ $p < 0.05$ および $p < 0.001$)。一方、シミュレーション記述子単独では統計的有意性に達しなかった。したがって Points Out 戦略では、シミュレーション記述子が改善傾向を示した ($\Delta R^2 = +0.034$)。ただし RMSE 改善テストによる評価では、この改善は統計的に有意ではなかった。

Mixtures Out 戦略では、全ての記述子条件で極めて有意な改善 ($p < 0.001$) が認められ、複合記述子で最大の効果 (係数 = 0.0342, $p < 0.001$) が観察された。また、モデリング記述子単独およびシミュレーション記述子単独もそれぞれ有意であり、両者が補完的に寄与することで最大の改善が得られることを示唆している。

Compounds Out 戦略における3つの記述子条件は統計的に有意な改善を示した ($p < 0.01$)。複合記述子 (モデリングとシミュレーション) が最も顕著な効果をもたらした (係数 = 0.0259, $p < 0.01$)。この結果は、未観測原料 (本研究ではポリオール) を含む製剤の予測において、追加記述子の組み込みが有益であることを示唆している。

表 5.4: 交差検証戦略ごとの特徴量強化効果の統計検定結果 (RMSE 改善).

CV Strategy	Feature Descriptors	Coefficient	Std. Error	p-value	Holm p-value	Significance
Points Out	Modeling descriptors	0.012	0.003	1.9e-04	3.9e-04	***
	Simulation descriptors	0.006	0.003	0.076	0.076	n.s.
	Modeling + Simulation descriptors	0.020	0.003	< 0.001	< 0.001	***
Mixtures Out	Modeling descriptors	0.028	0.005	< 0.001	< 0.001	***
	Simulation descriptors	0.018	0.004	< 0.001	< 0.001	***
	Modeling + Simulation descriptors	0.034	0.005	< 0.001	< 0.001	***
Compounds Out	Modeling descriptors	0.019	0.007	0.004	0.007	**
	Simulation descriptors	0.018	0.007	0.006	0.007	**
	Modeling + Simulation descriptors	0.026	0.008	0.001	0.003	**

Note: Modeling = Molecular modeling descriptors, Simulation = simulation descriptors, Modeling + Simulation = Combined descriptors.

Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s. = not significant.

All coefficients represent RMSE improvement (baseline - enhanced).

代表的な RMSE 結果として、複合記述子の平均 RMSE 改善値は Mixtures Out で約 0.034, Compounds Out で約 0.026, Points Out で約 0.020 であった。対照的に、Points Out におけるシミュレーション単独条件では RMSE が約 0.006 改善したが有意ではなかった。改善の程度は戦略によって異なり、Mixtures Out が最大の個別効果を示し、Compounds Out は有意ではあるがより小さな改善を示した。Points Out 戦略はより控えめな改善を示し、追加記述子の利点は文脈依存であることを示唆している。本研究の知見は、分子モデリングおよびシミュレーション記述子を機械学習モデルに組み込むことが、特に Compounds Out や Mixtures Out といった困難な一般化シナリオにおいて、洗浄率予測の予測性能を向上させるという統計的証拠を提供する。

また、界面活性剤相互作用の困難さにもかかわらず、Compounds Out 法は正の R^2 値を達成し、モデリングおよびシミュレーション記述子の組み込みにより精度が向上した (表 5.3 (a) および (b) 参照)。ただし、この知見はポリオール成分に限定されることに留意すべきである。これらの記述子が転移可能な分子情報を捕捉できることを示唆しており、それによって新規ポリオール原料組み合わせの予測を容易にする。

5.3.3 変数の影響分析

処方情報、モデリングパラメータ、シミュレーション記述子などの特徴量を追加したが、これらの特性値である洗浄率に影響を与える主要因を特定するため主成分分析 (PCA) を実施した (図 5.11)。

また、各記述子の寄与度を評価するため、表 5.3 (a) に示すように R^2 スコア 0.716 を達成したモデルに対して SHAP (SHapley Additive ExPlanations) 分析を実施した。この分析では原材料に関するデータを除外し、モデリングおよびシミュレーションパラメータにのみ焦点を当てている。SHAP 解析の結果を図 5.12 に示す。

SHAP 解析により、polyol beads と nonionic surfactant beads の比率が洗浄率予測において最も重要な因子であることが示された。この知見は、ポリオールが界面活性剤系の全体的な溶解性と安定性に寄与する一方、ノニオン性界面活性剤がミセルなどの高次構造形成や界面相互作用において重要な役割を果たすという基本的な理解と一致する。これらの成分は、製剤の自己組織化挙動に大きな影響を及ぼすことが実証されており、これが洗浄能力に影響を与える。

SHAP 解析において、シミュレーションパラメータである potential energy と pressure も上位にランクインした。この結果は、分子レベルの相互作用とシステム全体の熱力学的

安定性が性能の決定要因であることを示唆している。溶解度パラメータの差異は構造指標である一方、水以外の粒子数は組成指標である。これらの指標の AI への組み込みは、モデルの予測能力向上において共に重要であることを示している。

また、PCA 解析において、全分散の 67.0% を説明する第 1 主成分 (PC1) は、主に総界面活性剤含有量に関連する変動を捉え、非水成分に関連付けた。具体的には、PC1 に対して最も強い正の負荷量を示した変数は「Total number of non-water beads」であり、次いで「Potential energy」, 「Pressure」, 「Degree of amphiphilicity」が続いた。したがって、PC1 は界面活性剤および関連する非水成分の存在量を表すと解釈できる。PC1 スコアが高いほど、界面活性剤の総濃度が高く、関連する物理的特性への影響も大きいことを示す。第二主成分 (PC2) は追加で 15.5% の分散を説明し、親水性成分 (「Ratio of polyol beads」および「Fraction of hydrophilic beads」: 負の負荷量) と界面活性剤豊富成分 (「Ratio of nonionic surfactant beads」, 「Ratio of surfactant beads」, および関連する物理的特性, 正の負荷量) を明確に区別した。この明確な対立関係は、PC2 が主にシステム内のポリオール豊富な親水性領域と界面活性剤豊富な領域の組成バランスを表していることを示している。

さらに、SHAP 特徴重要度分析により「Ratio of polyol beads」と「Ratio of nonionic surfactant beads」が最も影響力の大きい特徴として特定された。この結果は、ポリオールとノニオン性界面活性剤成分の相対的バランスに対するシステムの極めて高い感受性を強調しており、PC2 によって特定された組成軸と密接に一致している。統計解析により、分子モデリングとシミュレーション記述子の組み込みが、様々な一般化シナリオにおいて予測精度を向上させることが示された。

総じて、PCA と SHAP は「どの成分・組み合わせが効くのか」を定量的に特定し、処方最適化における実務的な調整 (例: ポリオール対ノニオンのバランス) へと直結する知見を与える。これにより、配合反復の効率化が期待できる。

5.4 結論

機械学習モデルと分子モデリング・シミュレーション由来の記述子を統合し、洗顔フォーム製剤の予測フレームワークを開発した。本アプローチは、複雑な成分相互作用を有する多成分系における洗浄率予測の精度向上を目的とした。

評価戦略全体において、分子モデリングおよびシミュレーション由来の特徴量を取り入れることで、配合物のみのベースラインと比較して精度が向上した。Points Out 評価では、 R^2 値が 0.665 から 0.716 ($\Delta R^2 = +0.051$) に増加した。Mixtures Out 評価では 0.420 から 0.573 へ ($\Delta R^2 = +0.153$)、Compounds Out 評価では 0.023 から 0.341 へ ($\Delta R^2 = +0.318$) それぞれ向上した。回帰係数 (RMSE ベース OLS) の統計的有意性として、Mixtures Out ではモデリング記述子・シミュレーション記述子・複合記述子のいずれも有意な改善が確認され、Compounds Out でも同 3 条件で有意差が認められ、Points Out ではモデリング記述子と複合記述子で有意な改善が確認された。

これらのモデルは物理実験の代替とはならないが、実験スクリーニングの実用的なツールとして機能する。

今後の課題としては、(i) ポリオールに偏らない多様な界面活性剤を含むデータ拡張による汎化性能の向上、(ii) pH 依存性を明示的に組み込んだ分子間相互作用モデリング、(iii) 量子化学的記述子の導入と相互作用モデルの実験的検証による未知成分を含む配合への予測精度向上が挙げられる。

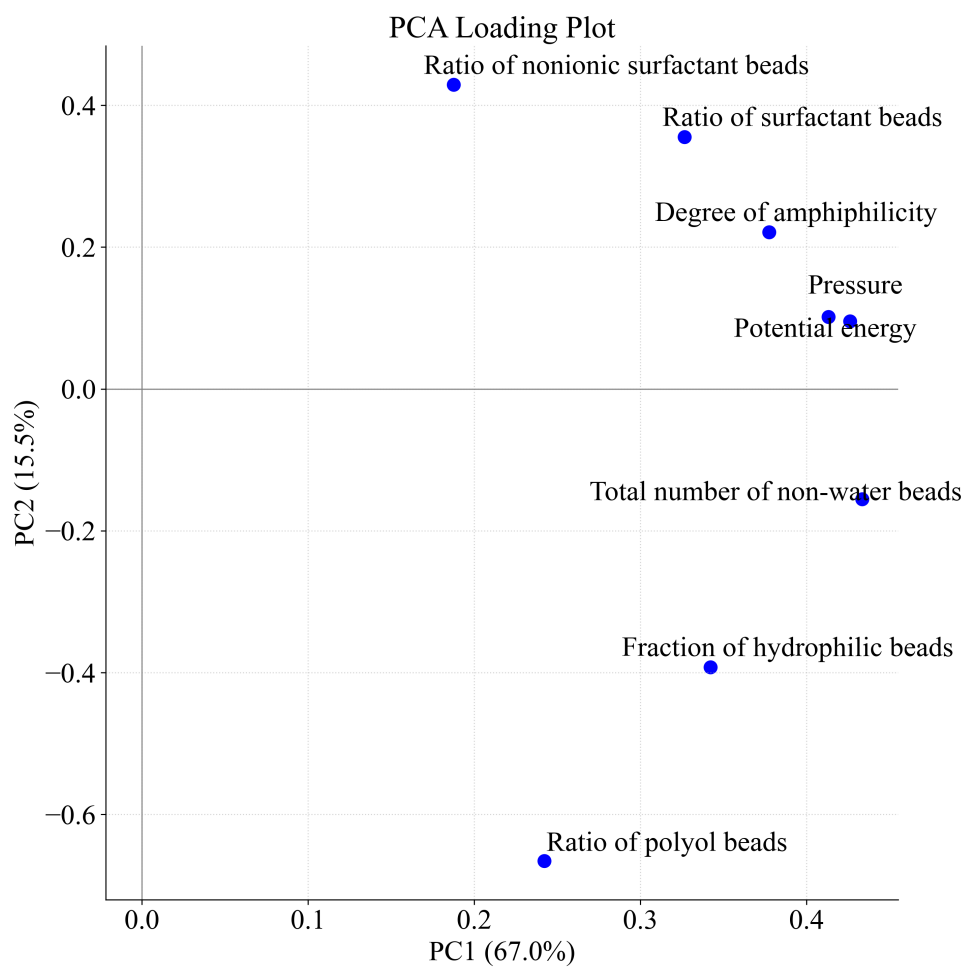


図 5.11: 選択特徴量の PCA ローディングプロット. PC1 (67.0%) は主に界面活性剤含有量と関連する物理的特徴による変動を表し, PC2 (15.5%) はポリオールとノニオン性界面活性剤成分の組成バランスを表す. 原料比率は除外した特徴群のみで解析.

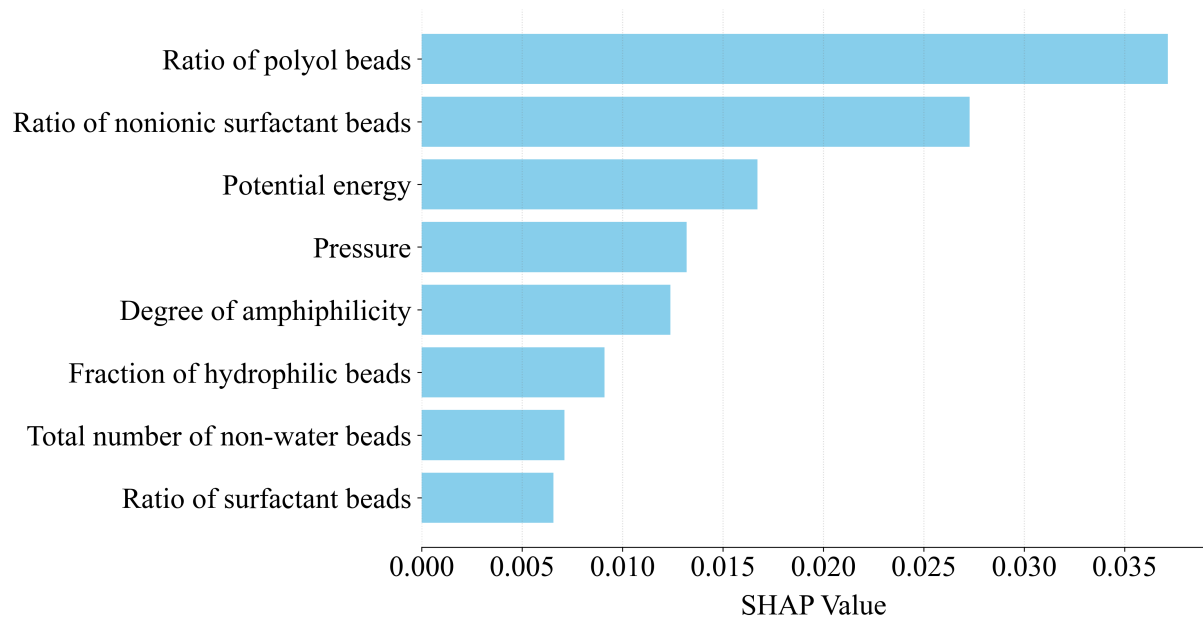


図 5.12: 原材料の組成データを除外し，モデリングおよびシミュレーション由来の記述子のみに焦点を当てた SHAP 解析結果．SHAP (SHapley Additive exPlanations) 値は各特徴がモデル予測に与える寄与を表す．原料比率は除外した特徴群のみで解析．

Chapter 6

結論，および，研究成果の応用と展望

6.1 研究目的に対する結論

本研究は、混合物を対象とする製品開発において、試作・実験を伴う反復による時間・費用の負担を軽減するため、AI技術(機械学習および深層学習)を活用した配合空間の事前スクリーニングと設計意思決定の迅速化・省実験化、ドメイン知識の獲得に資する枠組みの構築を目的とし、研究対象として錠剤とクレンジング剤の二系を設定した。錠剤ではAIのみで硬度・崩壊性・流動性・凝集性など複数特性値を同時に扱う多目的設計の実効性を示した。また、クレンジング剤では分子記述子、および、CAE (Computer Aided Engineering) 技術として粗視化モデルである Dissipative Particle Dynamics (DPD) を活用し、これらをAIを組み合わせることで、界面活性剤同士の複雑な相互作用の結果である洗浄率を予測する枠組みを提示した。第2章で設定した本論文の目的を以下に再掲する。

1. 少量データかつ外挿条件に対応する予測モデルと精度評価の枠組みを整備し、不確実性(信頼度)を定量化して提示する(第3,5章)。
2. 物性整合的な記述子体系として、粉体物性・粒度分布(第3章)、分子記述子(第4章)、およびDPD/粗視化モデリング由来の指標(第5章)を橋渡しして統合し、予測能力と説明性の両立を図る。
3. 実務スクリーニングに利用可能なモデル枠組みを構築し、候補配合の事前評価と意思決定を支援することで、試作の優先順位付けと知見の形式知化を促進する(第3章～第5章)。

目的1に対して、外挿領域を定量化した上でモデル精度を評価し、予測の信頼度を併せて提示する枠組みを整備した。具体的には、Applicability Domain (AD) による層別化(第3章)と、Points Out / Mixtures Out / Compounds Out の三段階による汎化レベル別の検証体系(第5章)を構築し、外挿リスクを把握しながら予測活用の可否を判断できるようにした。一方で、外挿領域の汎化は依然困難であることも明示した(第3,5章)。少量データ下でも外挿リスクを把握しながら予測活用を判断できる評価体系を構築した。

目的2に対して、粉体物性・粒度分布、分子記述子、およびDPD/粗視化モデリング由来の指標を統合し、混合物の状態を反映する物性整合的な記述子体系を構築した。粉体物性と粒度分布ベクトルの導入は予測精度向上に寄与し(第3章)、分子記述子は界面活性剤・多価アルコール等の構造情報を取り込む基礎となり、これも精度向上に寄与した(第4章)。さらに、DPD由来の状態量は、配合情報のみでは得られない非線形・相挙動の理解、ならびに外挿領域での予測精度向上に寄与すること、また、配合比のみで表現できない混合物挙動把握に寄与し、外挿領域下の予測向上にも寄与した(第5章)。

目的3に対して、配合候補を事前に評価して有望配合を抽出し、試作の優先順位付けと意思決定を支援する枠組みを構築した。データ統合と取得体制の整備により再学習可能なDBTL基盤を形成し（第3章）、多入力深層学習等のモデル設計や非線形学習の活用を通じて、実務スクリーニングに耐える精度を確保した（第3～5章）。また、SHAPやPCA、および逆解析に基づく層別解析により、寄与要因の可視化と説明性付与を行い、知見の形式知化に資する設計支援を実現した（第4,5章）。それらの結果として、スクリーニングとして活用される程度の精度を達成しつつ説明性を獲得した。

6.2 本研究の産業応用

本研究で構築した枠組みは、配合候補を予測モデルで事前評価し、実験に着手する前の早い段階で有望配合をふるい込むことで、試験回数を大幅に削減できるものである。これにより、人手による総当たり探索を置換・補完し、開発ライン全体の負荷軽減と開発速度の加速を両立できる。

また、クレンジング剤開発についても、洗浄率（洗浄性能）の予測精度を高めることで、試作の優先順位付けを半自動化し、開発期間短縮に直結させた。さらに、DPDによるシミュレーションと機械学習を組み合わせることで、「どの原料／原料組合せが効いているか」を特定し、その知見をそのまま設計指針へ落とし込める。

応用観点では、単純に実験削減を達成するだけでなく、過去到達できなかった特性値を得られ、パレートフロントを広げることができる。錠剤設計においては、原料情報を記述子として利用することで、外挿予測も積極的に活用し、第3章で構築したApplicability Domain評価により外挿性を定量的に確認しつつ、信頼できる範囲に限定して外挿予測を補助的に活用することで、試作の優先順位付けを早期に行い、試験回数の削減とリードタイム短縮を実現するとともに、パレートフロントの拡張と設計判断の一貫性向上に寄与している。

また、クレンジング剤開発においては、DPDはミセル化や相挙動といった自己組織化の可視化を可能にし、学習モデルの判断根拠を物理的に裏づけることで、いわゆるAIによるブラックボックス性を低減する役割を果たす。加えて、第4,5章で示した乱数処方作成に基づく逆解析による層別解析（特徴量効果の可視化）やSHAPに基づく重要特徴量の抽出とPCAによる共変動軸の可視化を実施し、各特徴量が予測性能や特性値に与える影響を定量的に抽出・比較した。これにより、現場開発者も気づかなかった個々の原料や組合せが寄与するメカニズムをモデル内部から説明可能にし、シミュレーションによる物理的知見と統合することで、より信頼性の高い設計支援を実現した。すなわち、本研究では「シミュレーションによる物理的裏づけ」と「逆解析による統計的裏づけ」の両面からブラックボックス化を回避し、解釈可能性を高めた点に特徴がある。

6.3 技術移転の方向性

本研究のワークフローは、混合物の特性値予測という観点において、配合情報と物理・化学記述子を統合して予測・提案を行う汎用枠組みであり、塗料、接着剤、乳化化粧品、食品、電池スラリーなど、他の混合物システムへ比較的容易に横展開できる。実際、本研究においては、錠剤とクレンジングという全く異なるシステムに対して、「混合物」という観点で類似の枠組みを適用した。もっとも、新たな系に適用する際には、当該系に応じた物性・分子記述子の設計と学習データの蓄積が前提となるが、本研究で構築したワークフローはそれらを効率的に取り込むための基盤として機能する。また、CAEの側面でも、DPDに限定されず、CFD・DEM・MD/CGMDなどのツールに置換可能な記述子層として設計しており、対象系に応じて計算レベルを選択できる。

これらのプロセスは「記述子生成 → 学習 → 提案」のモジュール化を徹底しているため、他ライン・他部門への短期に技術移転が可能である。さらに、既存の DoE（実験計画法）と重ね合わせ、モデルが示す有望領域に絞って試験設計を行うことで、試験効率を大きく高められる。LIMS/ELN など現場システムと接続すれば、データ収集から再学習までを一気通貫で回す運用にも適合する。

6.4 今後の研究展開

第3～5章の結果から、本研究の枠組みは内挿領域では一次スクリーニングに十分な精度を有する一方で、外挿性やデータの偏り、シミュレーションの物理的表現力に起因する限界が残されている。また、3章の錠剤開発の事例のように、実際の系では特性値ごとに取得データ量が不均一であることも多い。そのため、転移学習や半教師あり学習を取り入れ、データの偏りに頑健な学習を実現することで、少量データへの対応や未知原料に対する一般化性能をさらに強化する必要がある。加えて、物理制約付き学習や GCN (Graph Convolutional Networks) や PINN (Physics-Informed Neural Networks) を導入し、モデルの一貫性と外挿耐性を底上げすることを目指す。

また、シミュレーションという観点では、今回用いた DPD シミュレーション手法においては原料と汚れとの相互作用や、界面活性剤の電荷などの重要な物理現象が加味されていないため、MD も含めたより高度なシミュレーション手法の導入も検討が必要である。これらの発展的課題は、本論文で示した混合物設計のワークフローを、より広い系およびより過酷な外挿条件へ拡張するための次のステップであり、本研究はそのための基盤整備として位置づけられる。

謝辞

本博士論文は、筆者が慶應義塾大学大学院理工学研究科後期博士課程に在学中に、同大学理工学部機械工学科の荒井規允教授のご指導のもとで行った研究をまとめたものである。

主査の荒井規允教授には、共同研究を機に博士課程への進学を勧めていただき、私の人生の指針を大きく変えてくださったことに深く感謝申し上げます。先生には学術的な観点に留まらず、教育のあり方や周囲を巻き込むバイタリティ溢れる生き方に至るまで、一人の人間として深く感銘を受け、多くの知識と経験を学ばせていただきました。

学位審査にあたり、ご多忙中にもかかわらず副査の労を執っていただいた、泰岡顕治教授、深瀬康二教授、畑中美穂教授に対し、ここに厚く御礼申し上げます。

また、荒井研究室の皆様には多大なるご支援をいただきました。かつて在籍されていた佐藤さんや横山さんをはじめ、孟さん、山田さん、大里さんには、研究室の運営や論文執筆の作法など、多岐にわたる助言をいただきました。年齢の大きく離れる社会人の私に対しても、分け隔てなく接していただいたことは大きな支えとなりました。また、山口さん、按田さんの両氏には、同じ社会人ドクターという立場から、学術と企業の双方の視点に基づいた議論を通じ、多くの刺激をいただきました。研究室の皆様とのゼミや懇親会などの日々の交流は、私にとって何物にも代えがたい貴重な時間となりました。

株式会社ファンケルの三澤さんならびに足立さんには、私がAIについて未熟であった当初より共同プロジェクトに携わっていただいたこと、感謝の念に堪えません。お二人との二人三脚の協業がなければ、本論文の完成は到底叶いませんでした。本論文を提出するに至るまでの道程は、お二人との歩みがあってこそのものであります。

本研究の遂行にあたり、実験やデータ解析において技術的支援をいただいた研究室の皆様、ならびに多方面から支えてくださった方々に心より感謝いたします。

最後に、長きにわたる学業を応援してくれた父と母に感謝します。そして、仕事と学業の両立を献身的に支えてくれた妻・梓、ならびに娘の舞と息子の悠に、最大の感謝の意を表して謝辞といたします。

参考文献

- [1] Z. Li, L. Zhao, X. Lin, L. Shen, and Y. Feng. Direct compaction: An update of materials, trouble-shooting, and application. *Int. J. Pharm.*, Vol. 529, No. 1-2, pp. 543–556, 2017.
- [2] S. Shanmugam. Granulation techniques and technologies: recent progresses. *Bioim-pacts*, Vol. 5, No. 1, pp. 55–63, 2015.
- [3] I. Wünsch, I. Friesen, D. Puckhaber, T. Schlegel, and J. H. Finke. Scaling tabletting processes from compaction simulator to rotary presses—mind the sub-processes. *Pharmaceutics*, Vol. 12, No. 4, p. 310, 2020.
- [4] D. Markl and J. A. Zeitler. A review of disintegration mechanisms and measurement techniques. *Pharm. Res.*, Vol. 34, No. 5, pp. 890–917, 2017.
- [5] G. K. Bolhuis and Z. T. Chowhan. *Materials for direct compaction*, pp. 419–500. Marcel Dekker, New York, 2001.
- [6] V. Bhatia, A. Dhingra, B. Chopra, and K. Guarve. Co-processed excipients: Recent advances and future perspective. *J. Drug Delivery Sci. Technol.*, Vol. 71, p. 103316, 2022.
- [7] E. Jakubowska and N. Ciepluch. Blend segregation in tablets manufacturing and its effect on drug content uniformity—a review. *Pharmaceutics*, Vol. 13, No. 11, p. 1909, 2021.
- [8] R.M. Walters, G. Mao, E.T. Gunn, and S. Hornby. Cleansing formulations that respect skin barrier integrity. *Dermatol. Res. Pract.*, Vol. 2012, p. 495917, 2012.
- [9] Z. Li. Modern mild skin cleansing. *J. Cosmet. Dermatol. Sci. Appl.*, Vol. 10, No. 2, pp. 85–98, 2020.
- [10] D. Mijaljica, F. Spada, and I.P. Harrison. Skin cleansing without or with compromise: Soaps and syndets. *Molecules*, Vol. 27, No. 6, p. 2010, 2022.
- [11] T. Slavova, R. Stanimirova, K. Marinova, and K. Danov. Cleansing mechanisms and efficacy on artificial skin. *Molecules*, Vol. 30, No. 8, p. 1813, 2025.
- [12] Guy Broze, editor. *Handbook of Detergents, Part A: Properties*. CRC Press, 1999.
- [13] T.F. Tadros. *Applied Surfactants: Principles and Applications*. Wiley-VCH, 2005.

- [14] K. Watanabe, M. Masuda, K. Nakamura, T. Inaba, A. Noda, T. Yanagida, and T. Yanaki. A new makeup remover prepared with a system comprising dual continuous channels (bicontinuous phase) of silicone oil and water. *IFSCC Mag.*, Vol. 7, No. 4, pp. 310–318, 2004.
- [15] T. Iwanaga, K. Uchida, N. Takeuchi, and Y. Abe. Development of oil-type make-up remover prepared with polyglycerol fatty acid esters. *J. Soc. Cosmet. Chem. Japan*, Vol. 39, No. 3, pp. 186–194, 2005.
- [16] Wei Chen, Mei He, Li Xie, and Li Li. The optimal cleansing method for the removal of sunscreen: Water, cleanser or cleansing oil? *J. Cosmet. Dermatol.*, Vol. 19, No. 1, pp. 180–184, 2020.
- [17] P. Charoennit and N. Lourith. Validated UV-spectrophotometric method for the evaluation of the efficacy of makeup remover. *Int. J. Cosmet. Sci.*, Vol. 34, No. 2, pp. 190–192, 2012.
- [18] D.J. McClements. *Food Emulsions: Principles, Practices, and Techniques*. CRC Press, 2015.
- [19] Monzer Fanun, editor. *Microemulsions: Properties and Applications*. CRC Press, 2008.
- [20] T.F. Tadros. *Formulation Science and Technology*. De Gruyter, 2016.
- [21] M. P. Pileni, editor. *Structure and Reactivity in Reverse Micelles*. Elsevier, 1989.
- [22] M. Kahlweit, R. Strey, and D. Haase. Phase behavior of multicomponent systems water-oil-amphiphile-electrolyte. 3. *J. Phys. Chem.*, Vol. 89, No. 1, pp. 163–171, 1985.
- [23] J.N. Israelachvili. *Intermolecular and Surface Forces*. Academic Press, 2018.
- [24] Essential Chemical Industry. *Surfactants*, 2016.
- [25] M.J. Rosen and J.T. Kunjappu. *Surfactants and Interfacial Phenomena*. Wiley, 2012.
- [26] A.V. Rawlings and C.R. Harding. Moisturization and skin barrier function. *Dermatol. Ther.*, Vol. 17, No. Suppl 1, pp. 43–48, 2004.
- [27] Koichi Ishida, Akihiko Takahashi, Kotatsu Bito, Zoe Draelos, and Genji Imokawa. Treatment with synthetic pseudoceramide improves atopic skin, switching the ceramide profile to a healthy skin phenotype. *J. Invest. Dermatol.*, Vol. 140, No. 9, pp. 1762–1770.e8, 2020.
- [28] S. Yamaguchi, H. Li, and S. Imazato. Materials informatics for developing new restorative dental materials: a narrative review. *Front. Dent. Med.*, Vol. 4, p. 1123976, 2023.
- [29] Jinkui Wu, Shihui Wang, Li Zhou, Xu Ji, Yiyang Dai, Yagu Dang, and Markus Kraft. Deep-learning architecture in qspr modeling for the prediction of energy conversion efficiency of solar cells. *Ind. Eng. Chem. Res.*, Vol. 59, No. 42, pp. 18991–19000, 2020.

- [30] Arash Khajeh, Daniel Schweigert, Steven Torrisi, Linda Hung, Brian Storey, and Ha-Kyung Kwon. Early prediction of ion transport properties in solid polymer electrolytes using machine learning and system behavior-based descriptors of molecular dynamics simulations. *ChemRxiv*, 2023. Preprint.
- [31] J. Bajorath. Milestones in chemoinformatics: global view of the field. *J. Cheminform.*, Vol. 16, p. 124, 2024.
- [32] K. Funatsu. Process-informatics and development of polymer alloy. *Seikei-Kakou*, Vol. 34, No. 9, pp. 327–330, 2022.
- [33] G Absalan, B Hemmateenejad, M Soleimani, M Akhond, and R Miri. Quantitative structure–micellization relationship study of gemini surfactants using genetic-PLS and genetic-MLR. *QSAR Comb. Sci.*, Vol. 23, No. 6, pp. 416–425, 2004.
- [34] Shadpour Mallakpour, Mehdi Hatami, and Hassan Golmohammadi. Prediction of inherent viscosity for polymers containing natural amino acids from the theoretical derived molecular descriptors. *Polymer*, Vol. 51, No. 15, pp. 3568–3574, 2010.
- [35] Pathan Mohsin Khan, Bakhtiyor Rasulev, and Kunal Roy. QSPR modeling of the refractive index for diverse polymers using 2D descriptors. *ACS Omega*, Vol. 3, No. 10, pp. 13374–13386, 2018.
- [36] Théophile Gaudin, Patricia Rotureau, and Guillaume Fayet. Mixture descriptors toward the development of quantitative Structure–Property relationship models for the flash points of organic mixtures. *Ind. Eng. Chem. Res.*, Vol. 54, No. 25, pp. 6596–6604, 2015.
- [37] Mohammad Amin Sobati, Danial Abooali, Babak Maghbooli, and Hamidreza Najafi. A new structure-based model for estimation of true critical volume of multi-component mixtures. *Chemom. Intell. Lab. Syst.*, Vol. 155, pp. 109–119, 2016.
- [38] Bishal Das, Sayanee Sarkar, Sourav Paul, Ayesha Zafar, Deepika Biswas, Jigisha Roy Panda, and Tamalika Chakraborty. Mitigation strategies of emerging pollutants from cosmetic industry effluent: A review. *Eur. J. Theor. Appl. Sci.*, Vol. 2, No. 3, pp. 886–899, 2024.
- [39] N A Vita, C A Brohem, A D P M Canavez, C F S Oliveira, O Kruger, M Lorencini, and C M Carvalho. Parameters for assessing the aquatic environmental impact of cosmetic products. *Toxicol. Lett.*, Vol. 287, pp. 70–82, 2018.
- [40] Mainak Chatterjee and Kunal Roy. Prediction of aquatic toxicity of chemical mixtures by the QSAR approach using 2D structural descriptors. *J. Hazard. Mater.*, Vol. 408, p. 124936, 2021.
- [41] Yoshihiro Hayashi, Yuki Marumo, Takumi Takahashi, Yuri Nakano, Atsushi Kosugi, Shungo Kumada, Daijiro Hirai, Kozo Takayama, and Yoshinori Onuki. In silico predictions of tablet density using a quantitative structure-property relationship model. *Int. J. Pharm.*, Vol. 558, pp. 351–356, 2019.
- [42] Alireza Abbasi and Reza Eslamloueyan. Determination of binary diffusion coefficients of hydrocarbon mixtures using MLP and ANFIS networks based on QSPR method. *Chemom. Intell. Lab. Syst.*, Vol. 132, pp. 39–51, 2014.

- [43] Eugene N Muratov, Ekaterina V Varlamova, Anatoly G Artemenko, Pavel G Polishchuk, and Victor E Kuz'min. Existing and developing approaches for QSAR analysis of mixtures. *Mol. Inform.*, Vol. 31, No. 3-4, pp. 202–221, 2012.
- [44] Guillaume Fayet and Patricia Rotureau. QSPR models to predict the physical hazards of mixtures: a state of art. *SAR QSAR Environ. Res.*, Vol. 34, No. 9, pp. 745–764, 2023.
- [45] Christoforos Brozos, Jan G. Rittig, Elie Akanny, Sandip Bhattacharya, Christina Kohlmann, and Alexander Mitsos. Predicting the temperature-dependent CMC of surfactant mixtures with graph neural networks. arXiv, 2024. Preprint.
- [46] Yick Eu Chew, Ho Yan Lee, Yi Pei Heng, Siang Tee Tiew, Jia Wen Chong, and Nishanth G. Chemmangattuvalappil. Design of formulated fragrant products using rough set machine learning and molecular design tools. *Chem. Eng. Res. Des.*, Vol. 203, No. C, pp. 305–320, 2024.
- [47] Zohreh Faramarzi, Fatemeh Abbasitabar, Vahid Zare-Shahabadi, and Hossein Jalali Jahromi. Novel mixture descriptors for the development of quantitative structure – property relationship models for the boiling points of binary azeotropic mixtures. *J. Mol. Liq.*, Vol. 296, p. 111854, 2019.
- [48] Ting Wang, Lili Tang, Feng Luan, and M. Natália D. S. Cordeiro. Prediction of the toxicity of binary mixtures by QSAR approach using the hypothetical descriptors. *Int. J. Mol. Sci.*, Vol. 19, No. 11, p. 2984, 2018.
- [49] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Adv. Neural Inf. Process. Syst.*, pp. 4765–4774, 2017.
- [50] Yi Wang, Bo Zhang, Qiang Li, and Sheng Chen. Molsets: High-throughput graph neural networks for multicomponent mixture property prediction. *Nat. Commun.*, Vol. 14, p. 5123, 2023.
- [51] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. arXiv, 2020. Preprint.
- [52] RDKit: Open-source cheminformatics software. Accessed: 2023-12-31.
- [53] Chun Wei Yap. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.*, Vol. 32, No. 7, pp. 1466–1474, 2011.
- [54] D. P. Faasen, W. G. Bouwman, J. C. Schouten, and J. M. Kleijn. Hansen solubility parameters obtained via molecular dynamics simulations as a route to predict siloxane surfactant adsorption. *J. Colloid Interface Sci.*, Vol. 575, pp. 326–336, 2020.
- [55] O. Afzal, A. S. A. Altamimi, M. S. Nadeem, S. I. Alzarea, W. H. Almalki, S. Ahmad, A. S. Alfawaz Altamimi, M. A. Alossaimi, S. Alshehri, M. M. Ghoneim, and I. Kazmi. Hansen solubility parameters and green nanocarrier based removal of trimethoprim from contaminated aqueous solution. *J. Mol. Liq.*, Vol. 357, p. 119657, 2022.
- [56] Miron B. Kursá, Aleksander Jankowski, and Witold R. Rudnicki. Boruta – a system for feature selection. *Fundamenta Informaticae*, Vol. 101, No. 4, pp. 271–285, 2010.

- [57] P. J. Hoogerbrugge and J. M. V. A. Koelman. Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *Europhys. Lett.*, Vol. 19, No. 3, p. 155, 1992.
- [58] P. Español and P. Warren. Statistical mechanics of dissipative particle dynamics. *Europhys. Lett.*, Vol. 30, No. 4, p. 191, 1995.
- [59] Robert D. Groot and Patrick B. Warren. Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *J. Chem. Phys.*, Vol. 107, No. 11, pp. 4423–4435, 1997.
- [60] Igor V. Pivkin and George E. Karniadakis. Coarse-graining limits in open and wall-bounded dissipative particle dynamics systems. *J. Chem. Phys.*, Vol. 124, No. 18, p. 184101, 2006.
- [61] J. H. Hildebrand and R. L. Scott. *The Solubility of Nonelectrolytes*. ACS monograph. Reinhold Publishing Corporation, 1950.
- [62] Amitesh Maiti and Simon McGrother. Bead–bead interaction parameters in dissipative particle dynamics: Relation to bead-size, solubility parameter, and surface tension. *J. Chem. Phys.*, Vol. 120, No. 3, pp. 1594–1601, 2004.
- [63] JSOL Corporation. J-octa: Integrated simulation system for soft materials, 2023. Version 5.0 [Computer software].
- [64] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, Vol. 25, No. 9, pp. 1157–1174, 2004.
- [65] Yiyong Mai and Adi Eisenberg. Self-assembly of block copolymers. *Chem. Soc. Rev.*, Vol. 41, No. 18, pp. 5969–5985, 2012.
- [66] Nathaniel A. Lynd, Adam J. Meuler, and Marc A. Hillmyer. Polydispersity and block copolymer self-assembly. *Prog. Polym. Sci.*, Vol. 33, No. 9, pp. 875–893, 2008.
- [67] Yuuki Ishiwatari, Takahiro Yokoyama, Tomoya Kojima, Taisuke Banno, and Noriyoshi Arai. Machine learning prediction of self-assembly and analysis of molecular structure dependence on the critical packing parameter. *Mol. Syst. Des. Eng.*, Vol. 9, No. 1, pp. 20–28, 2024.