

厚生労働行政推進調査事業費補助金（化学物質リスク研究事業）  
In silico予測手法の高度化とNew Approach Methodologyの活用に基づく化学物質の統合的ヒ  
ト健康リスク評価系の基盤構築に関する研究

令和5年度 分担研究報告書

機械学習を用いた皮膚感作性試験代替法の開発と  
化学物質のリスク評価への活用に関する研究

研究分担者 安部 賀央里

名古屋市立大学大学院薬学研究科 レギュラトリーサイエンス分野 講師

研究協力者 足利 太可雄

国立医薬品食品衛生研究所 安全性予測評価部 室長

研究協力者 頭金 正博

名古屋市立大学大学院薬学研究科 レギュラトリーサイエンス分野

研究協力者 木下 啓

名古屋市立大学大学院薬学研究科 レギュラトリーサイエンス分野

研究協力者 徳永 朱莉

名古屋市立大学大学院薬学研究科 レギュラトリーサイエンス分野

**研究要旨**

動物実験の規制に伴い、皮膚感作性評価における動物実験代替法の開発が急務となっている。2021年には皮膚感作性を評価する際のDefined ApproachesがOECDのガイドラインNo.497に公表され、有害性評価や感作性強度評価において*in chemico*、*in vitro*、*in silico*を組み合わせた考え方が示されている。本課題ではガイドラインNo.497にて検討されたマウスの局所リンパ節試験における実験結果（皮膚感作性の強度指標）を学習データとして、OECDのテストガイドラインに記載されている*in vitro*試験結果やOECD QSAR Toolboxの化学物質情報を組み合わせて機械学習による皮膚感作性強度予測モデルを構築した。開発した予測モデルを基に、性能評価手法や適用領域の検討を行い、規制安全性評価への受け入れを目指した実用的な*in silico*モデル開発における課題を検討した。また、テストガイドラインに記載されている*in vitro*試験法について、化学構造情報等から実験結果を予測する*in silico*モデルの開発を行い、組み合わせ評価における有用性を検証した。さらに、化学物質の構造情報からヒトでの皮膚感作性の有無を判別する高性能かつ汎化性の高い予測モデルの構築も行った。

## A. 研究目的

皮膚感作性物質は、繰り返しの暴露によって皮膚に局所的な炎症(かぶれ)が生じるアレルギー性接触皮膚炎(ACD)を引き起こす化学物質であり、T細胞を介したIV型過敏性免疫反応を誘発すると考えられている。ACDは罹患者の多い疾患であり、生活の質に大きな影響を与えるため、皮膚感作性を引き起こす可能性のある化学物質を適切に評価し、規制することは極めて重要である。これまで、化学物質の皮膚感作性の評価のためのヒトおよび動物を使用する試験法が数多く開発されてきた。特に、マウス局所リンパ節アッセイ(LLNA)は、動物への苦痛が少なく、定量的な評価が可能なことから国際的に推奨されている。しかし、欧州で2013年以降、動物実験が実施された成分を含む化粧品の販売が禁止されたため、皮膚感作評価における動物実験代替法の開発が早急に求められている。また、動物実験にかかる費用、時間、動物愛護の観点から、化粧品業界のみならず、世界中の様々な業界において動物を用いない動物実験代替法の開発が加速している。

これまでに、皮膚感作の有害性発現経路(AOP)の解明が進んでおり、2012年にOECDガイダンス文書として公表されている。皮膚感作性の動物実験代替法として、AOPにおける、各Key Eventに対応した*in chemico*や*in vitro*手法が開発され、OECDのテストガイドラインに記載されている。一方で、皮膚感作性は免疫反応が関与する複雑なメカニズムであるため、複数の試験法を組み合わせる総合的に評価するIntegrated Approach to Testing and Assessment(IATA)が提唱されており、複数の*in vitro*試験結果や*in silico*の結果を組み合わせ

評価を行う統合的アプローチが検討されている。2021年6月には皮膚感作性を評価する際のDefined ApproachesがOECDのガイドラインNo.497に公表され、有害性評価や感作性強度評価において*in chemico*、*in vitro*、*in silico*を組み合わせた考え方が示されている。

人工知能技術の一つである機械学習は、化学物質の毒性評価における*in silico*手法の開発において重要な役割を示し、多種多様な毒性関連データを有効活用する手段として大きな期待が集まっている。我々は、機械学習を用いて、皮膚感作性の強度指標であるLLNAのEstimated Concentration needed to produce a stimulation index of 3(EC3)値を予測する回帰モデルの構築に取り組んできた(Ambe *et al.*, 2021)。

本課題では、皮膚感作性に関する信頼性の高い*in vivo*や*in vitro*の実験結果を収集し、AOPの主要キーイベントを評価する*in vitro*の実験結果や物性値等の化学物質情報から*in vivo*での皮膚感作性を予測する機械学習モデルを構築する。既存の*in silico*モデルとの比較検討を行い、モデルの性能や適用領域について評価する。また、開発した予測モデルの適用領域の明確化や拡大に向けて、キーイベントを評価する*in vitro*実験結果の*in silico*モデル化や、三次元の画像情報等を使用した予測モデルの開発にも取り組む。

さらに、他の毒性エンドポイントへの適用を視野に入れ、開発した予測モデルを化学物質のリスク評価に活用するにあたり、具体的な導入場面を想定したケーススタディを行い、予測結果に求められる信頼性や予測性能を検討し、安全性評価規制への受け入れへ向けた具体的な課題を整理する。

## B. 研究方法

## B.1 皮膚感作性の強度指標予測モデルの改良

本課題ではOECDの皮膚感作性に関する Defined Approachのガイドライン (No.497) において検討された化学物質の試験情報を使用してモデルの改良を行った。皮膚感作性の*in silico*予測モデルの精緻化として、強感作性物質の予測を重視した回帰モデルを構築や適用領域を設定した。Defined ApproachのOECDガイドライン (No.497) において検討された化学物質の試験情報を使用した。皮膚感作性の強度指標であるLLNAのEC3値を目的変数とし、皮膚感作性のAOPに関する*in vitro*試験結果 (DPRA、KeratinoSens<sup>TM</sup>、h-CLAT)、物性値、OECD QSAR ToolBoxから得られる化学情報を説明変数とし、勾配ブースティング決定木系アルゴリズムを用いた回帰モデルを構築した。本モデルを用いて、2023年に公表されたOECDのQSAR Assessment Frameworkに関するガイダンスを基に、(Q)SARモデル評価項目について検討した。

## B.2 *In vitro* 実験の予測モデル構築

皮膚感作性のAOPに関する*in vitro*実験結果を*in silico*で予測することは、難溶性物質等の*in vitro*実験の実施が困難な化学物質への対応など欠損値補完に対する有益なアプローチとなる。本課題では、皮膚感作性のAOPにおいて分子開始イベントMIE (タンパク質との共有結合) を評価する比較的簡便な*in chemico*試験であるDPRA (OECD TG442C) とKey Event 2 (角化細胞における炎症性応答) を評価するKeratinoSens<sup>TM</sup> (OECD TG442D) に着目し、DPRAの実験結果とMOEで計算した分子記述子情報のみからKeratinoSens<sup>TM</sup>の実験結果を予測する*in*

*silico*モデルを構築した。勾配ブースティング決定木系のアルゴリズムXGBoostを使用して機械学習による判別モデルを構築した。Hoffmannらの文献のthe Cosmetics Europe database (*Crit Rev Toxicol.* 2018; 48(5): 344-358) に含まれる物質をモデル構築用の学習データとした。学習データは陽性物質と陰性物質の割合を保持したまま層化分割し、訓練データと検証データとした。訓練データを用いてパラメータ調整を行い、検証データの結果から最適なモデルを構築した。また、モデル作成に用いたデータソースとは異なるUrbischのデータセット (*Regul Toxicol Pharmacol.* 2015; 71(2): 337-351) から外部検証データを作成し、構築した予測モデルの汎用性を確認した。説明変数としてDPRAの実験値、MOEで計算した分子記述子を使用した。

さらに、KeratinoSens<sup>TM</sup>の*in silico*モデルの有用性を確認するために、Defined Approachのガイドライン (No.497) にて示されたハザード同定のためのDefined Approachである”2 out of 3” (2試験の一致した結果に基づき評価する方法) において、KeratinoSens<sup>TM</sup>の実測値を使用した場合と本モデルによる予測値を使用した場合の組み合わせ評価の性能を比較した。

## B.3 ヒトの皮膚感作性評価の予測モデル構築

ヒトの皮膚感作性評価を化学物質の構造情報のみから*in silico*で予測する機械学習モデルを構築した。Goldenらの文献 (*ALTEX.* 2021; 38(1): 33-48) に加えてHoffmannらの文献 (*Regul Toxicol Pharmacol.* 2022; 131: 105169)、Basketterらの文献 (*Dermatitis.* 2014; 25(1): 11-21)、Apiらの文献 (*Dermatitis.*

2017; 28(5): 299-307) に記載されているヒトでの皮膚感作性に関する情報を学習データとして利用した。目的変数をヒトの皮膚感作性の有無、説明変数を化学構造情報である分子記述子とし、機械学習アルゴリズムとして勾配ブースティング決定木であるCatboostを用いて判別モデルを構築した。分子記述子はMOEにより計算した。また、外部検証データは、信頼性が高く情報量の多い、Hoffmann (2022)、Basketter (2014)、Api (2017)の3つの文献から陽性20物質、陰性20物質の計40物質をランダムに抽出し、構築した予測モデルの汎用性を確認した。また、*in silico*モデルにおいて重要なデータ分割、性能評価、変数重要度によるモデルの説明性、適用領域について検討した。

(倫理面への配慮) 本研究は動物を用いた研究を行わないため対象外である。

## C. 研究結果

### C.1 皮膚感作性の強度指標予測モデルの改良

これまでに我々が開発した皮膚感作性の強度指標 (LLNA EC3値) の予測モデルを基に、Defined Approach のガイドライン (No.497) において検討された化学物質の試験情報を使用してモデルの改良を行った。LLNA EC3値情報が得られた154物質を用いて、XGBoost 1.7.3 (python 3.8.13)による回帰モデルを構築し、層化5分割交差検証による内部検証を実施した。モデルの適用領域判定は、説明変数を使用したk近傍法 (図1) からデータ密度の高い領域を設定する方法と、皮膚感作性に関する*in vitro*試験法 (DPRA, KeratinoSens<sup>TM</sup>, h-CLAT) の実験結果を用いて設定する方法を組み合わせた。Hoffmann

(2022)の72物質を用いて外部検証も実施した。内部検証では決定係数R<sup>2</sup>値が0.67であった。また、GHS区分による3段階の感作性強度カテゴリーに予測EC3値を分類したところ、適用領域内物質における一致率は0.79だった。外部検証の適用領域内物質におけるGHS区分の一致率は0.50であった。

本モデルの規制評価での使用にむけた課題について、OECDのQSAR Assessment Frameworkに関するガイダンスを基に、*in silico*予測モデル評価について報告書式 (QSAR model reporting format) に従って検討した。公表されている信頼性の高い学習データや算出方法が明確な説明変数、予測モデルの透明性については概ね問題ないことが確認された。一方で、適用領域や外部検証の妥当性については、ガイダンス自体の基準も解釈が難しく、引き続き検討する必要がある。

### C.2 *In vitro* 実験の予測モデル構築

KeratinoSens<sup>TM</sup>の予測モデル化では、Hoffmann (2018)らの文献から122物質を訓練用データセット、Urbisch (2015)らの文献から102物質を検証用データセットとした。MOEを用いて算出した分子記述子とDPRAの実験値2種 (システインペプチド減少率とリジンペプチド減少率) の計444種から選択された7種の記述子を説明変数とした。XGBoost 1.4.2 (python 3.8.12)による分類モデル (KeratinoSens<sup>TM</sup>の陽性、陰性を判別) は、検証用データセットにおいて、正確度0.80、ROC-AUC 0.84の高い予測性能を示した。また、構築した判別モデルの変数重要度を確認したところ、最も重要な変数はDPRAの実験値であるシステインペプチド減少率であった。

Defined Approachである”2 out of 3”の組み合わせ評価において、使用するKeratinoSens™の結果を実験値の場合と、我々のモデルによる予測値の場合で比較した。検証用データセットに用いたUrbisch (2015)らの文献から48物質を用いた。”2 out of 3”による感作性判定結果と、LLNAにおける感作性の有無の一致率を確認した。KeratinoSens™の実験値を使用した”2 out of 3”では、LLNAの結果との一致率0.73、感度0.85、特異度0.47だった。一方で、KeratinoSens™の*in silico*予測値を使用した場合は、一致率0.77、感度0.82、特異度0.67だった。KeratinoSens™の予測値（陽性、陰性の二値）は実験値と比較して、LLNAとの結果の一致率、感度が高い傾向が確認できた。以上の結果からKeratinoSens™の予測モデルは組み合わせ評価において有用である可能性が示唆された。

### C.3 ヒトの皮膚感作性評価の予測モデル構築

ヒトの皮膚感作性予測モデルにおいて、Golden(2020)らの文献に加えて、Hoffmann (2022)、Basketter (2014)、Api (2017)らの文献情報を追加し、陽性293物質、陰性182物質を学習用データセットとした。また、これまでの検証用データセットは陽性物質と陰性物質の不均衡が大きかったため、Hoffmann (2022)、Basketter (2014)、Api (2017)の3つの信頼性が高い文献情報から40物質（陽性20物質、陰性20物質）を検証用データセットとした。MOEを用いて算出した440種の分子記述子から変数重要度の高い134種を説明変数とし、ヒトの感作性の有無を目的変数とした。検証用データセットにおいて、感度0.85、特異度0.65、ROC-AUC 0.88であり、高

い予測性能を示した。分類モデルはPython3.8.5のCatBoostを使用した。また、変数重要度の高い134種において、LUMOに関する3つの変数は全て含まれていた。本モデルの適用領域はモデル構築に使用した分子記述子から計算したユークリッド距離を用いて、k近傍法 (k=5) により設定したデータ距離内の物質とした (図1)。

### D. 考察

皮膚感作性強度を予測する機械学習モデルでは、OECDの皮膚感作性に関するDefined Approachのガイドライン (No.497) において公開された化学物質の試験情報を使用し、強度指標であるLLNAのEC3値を高性能で予測することができた。本モデルはOECDが公開した信頼性の高いデータを用いたことや、テストガイドラインに掲載された*in vitro*試験法の結果、OECD QSAR ToolBoxを用いた化学物質情報を用いており、予測性能と解釈性・透明性を有している。一方で、本モデルを使用した外部検証の実施や適用領域を設定したが、これらの妥当性については明確な基準の整備が十分ではないため、さらなる検討が必要であることが明らかとなった。

*In vitro*実験値の予測モデル化において、KeratinoSens™の結果を予測する*in silico*モデルは既存の文献情報を有効活用し、高い予測性能と汎化性能を示した。*In vitro*実験は動物を使用しない代替法として有用だが、化学物質の入手や、*in vitro*実験において問題となる溶解性などの観点から、実験結果が得られない場合がある。*In silico*での高性能な*in vitro*実験値予測モデルの開発は、皮膚感作性の評価において実験値が無い場合の穴埋めや、*in vitro*実験が必要かどうかの

スクリーニングにも使用できる。また、Defined Approachである”2 out of 3”の組み合わせ評価において、本モデルのKeratinoSens™予測値は、実験値を使用した場合と比較して同等またはそれ以上の予測性能(LLNAにおける皮膚感作性の有無との一致)が確認されたため、代替法の組み合わせによる皮膚感作性評価において一定の有用性が期待できる。

ヒトの皮膚感作性予測モデルにおいて、化学物質の構造情報のみから、感作性の有無を精度良く判別する機械学習モデルが構築できた。また、本モデルでは、*in silico*モデルにおいて重要なデータ分割、性能評価手法、変数重要度によるモデルの説明性、適用領域についても検討しており、QSARモデル構築へのフィードバックが可能となる。

## E. 結論

機械学習による皮膚感作性の予測モデル構築については、信頼性の高い学習データを用いて高性能で感作性強度を予測する回帰モデルを構築した。本モデルに対し、OECD QSAR Assessment Frameworkガイダンスに基づいた(Q)SARモデル評価項目について検討し、適用領域や外部検証の妥当性については課題点が多いことが確認された。また、皮膚感作性のキーイベントである*in vitro*実験結果を*in silico*で予測する高性能かつ汎化性能の高いモデルを構築し、Defined Approachへの適用を検証した。化学物質の構造情報のみからヒトでの皮膚感作性の有無を判別する高性能なモデルも構築し、性能評価や適用領域も検討した。

本課題における皮膚感作性を対象とした*in silico*予測モデルの開発やリスク評価への活用に関する課題検討の結果は、他の毒性

エンドポイントに対して*in silico*アプローチを適用する場合においても共通する検討事項として参考となることが期待される。

## F. 研究発表

### F.1. 論文発表

1. Ashikaga T, Narita K, Kobayashi M, Tachibana S, Murasaki W, Suzuki M, Ambe K, Tohkin M. Skin sensitization potency prediction of ingredients in hair colorants using *in silico* models of machine learning. *Journal of Japanese Cosmetic Science Society*. **47**(1), 1-5, 2023.
2. 安部賀央里. 皮膚感作性強度の予測に向けた機械学習による *in silico* アプローチ. *フレグランスジャーナル*, **7**, 16-20, 2023
3. 安部賀央里. 機械学習を用いた皮膚感作性試験代替法の開発. *コスメティックステージ*, 2月号, 60-65, 2023.
4. 安部賀央里. 【第V編 化粧品業界】第4章 皮膚感作性 *in silico* 予測モデル (動物実験代替法と New Approach Methods の開発・利用動向. 監修: 小島肇夫. 発行日: 2023.9.29, シーエムシー出版)

### F.2 学会発表

1. Linfeng Liu, Mayu Onishi, Yuka Yoshii, Kaori Ambe, Toshiaki Makino, Masahiro Tohkin: Efficacy and safety of new class disease-modifying anti-rheumatic drugs: Bayesian and frequentist network meta-analysis. 日本薬学会第144年会 (2024.3.28 – 3.31, 横浜)
2. 安部賀央里, 木下 啓, 山田 隆志, 足利 太可雄, 頭金 正博: 皮膚感作性評

- 価における *in silico* 予測モデルの開発. 第 10 回 CBI 学会個別化医療研究会 (2024.2.27, 岐阜)
3. 青木 優佳, 安部 賀央里, 頭金 正博, 村島 美穂, 濱野 高行, 和知野 千春, 木村 和哲, 日比 陽子, 近藤 勝弘: 電子カルテ情報と機械学習を活用したシスプラチン誘発性急性腎障害の予測モデルの構築. 第 44 回日本臨床薬理学会学術総会 (2023.12.14 - 12.16, 神戸)
  4. Linfeng Liu, Mayu Ohnishi, Yuka Yoshii, Kaori Ambe, Masahiro Tohkin: Comparison of the efficacy and safety of various anti-rheumatic drugs for rheumatoid arthritis: Network Meta-Analysis. 第 44 回日本臨床薬理学会学術総会 (2023.12.14 - 12.16, 神戸)
  5. 徳永 朱莉, 安部 賀央里, 山田 隆志, 足利 太可雄, 頭金 正博: 過小評価の回避を指向した皮膚感作性強度を予測する機械学習モデルの開発. 日本動物実験代替法学会 第 36 回大会 (2023.11.27 - 11.29, 千葉)
  6. 青木 優佳, 安部 賀央里, 頭金 正博, 村島 美穂, 濱野 高行: 悪性腫瘍患者における化学療法中の低 Na 血症を予測する機械学習モデルの構築およびリスク因子の検討. 第 33 回日本医療薬学会年会 (2023.11.3 - 11.5, 仙台)
  7. 大西 真由, 劉 臨風, 吉井 優花, 安部 賀央里, 頭金 正博: ネットワークメタアナリシスによる新規抗リウマチ薬の有効性と安全性の比較検討. 第 33 回日本医療薬学会年会 (2023.11.3 - 11.5, 仙台)
  8. 木下啓, 安部 賀央里, 山田 隆志, 足利 太可雄, 頭金 正博: Establishment of *in silico* prediction model for skin sensitization aiming for practical application. CBI 学会 2024 年大会 (2024.10.23 - 10.26, 東京)
  9. Tamana Enami, Kaori Ambe, Masahiro Tohkin: Development of a model for predicting a chemical-induced CYP inhibitory activity using images data showing chemical structure information. 2023 年 ICCP450/JSSX 国際合同大会 (2023.9.25 - 9.29, 静岡)
  10. 出来 佑都, 安部 賀央里, 頭金 正博, 和知野 千春, 木村 和哲, 日比 陽子, 近藤 勝弘, 村島 美穂, 濱野 高行: 電子カルテ情報と機械学習を用いたバンコマイシン誘発性急性腎障害の予測手法開発. 第 9 回次世代を担う若手のためのレギュラトリーサイエンスフォーラム (2023.9.16, 東京)
  11. 濱上 敦史, 土井 更良, 安部 賀央里, 頭金 正博: 副作用報告データベースとトランスポーター阻害活性情報を用いた機械学習による胆汁うっ滞型薬物性肝障害の予測モデルの開発. 第 69 回日本薬学会東海支部大会 (2023.7.8, 名古屋)
  12. 足利 太可雄, 波多野 浩太, 岩佐 帆乃夏, 木下 啓, 中村 伸昭, 安部 賀央里, 頭金 正博: Next Generation Risk Assessment 事例研究: ヘアカラー製剤に存在するバンドロフスキーベースの定量的皮膚感作性リスク評価. 第 48 回日本化粧品学会 (2023.6.23 - 6.24, 東京-Online Hybrid)
  13. 安部 賀央里: データベースと機械学習を活用した毒性予測へのアプローチ～有機化学から視野を広げて～ シン

- ポジウム29 トキシコロジストのキャリア形成支援プログラム：デジタル時代の人材育成と教育. 第50回日本毒性学会学術年会（2023.6.19-6.21, 横浜）
14. 木下 啓, 安部 賀央里, 山田 隆志, 足利 太可雄, 頭金 正博: 機械学習を用いた皮膚感作性強度予測モデルの開発と実用化に向けた検討. 第50回日本毒性学会学術年会（2023.6.19-6.21, 横浜）
15. Linfeng Liu, Mayu Ohnishi, Yuka Yoshii, Kaori Ambe, Masahiro Tohkin: Efficacy and Safety of Anti-rheumatic Drugs in Patients with Rheumatoid Arthritis: Bayesian Network Meta-Analysis. 第50回日本毒性学会学術年会（2023.6.19-6.21, 横浜）
16. 大西 真由, 劉 臨風, 吉井 優花, 安部 賀央里, 頭金 正博: 新規抗リウマチ薬の有効性と安全性に関する研究—ネットワークメタ解析手法を用いた検討—. 第7回日本臨床薬理学会東海北陸地方会（2023.6.4, オンライン）

G. 知的財産権の出願・登録状況  
（予定を含む。）

1. 特許取得  
なし
2. 実用新案登録  
なし
3. その他  
なし

**K近傍法:**

近傍k個のデータを用いてデータ密度を推定する方法

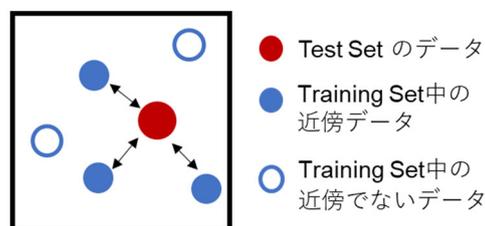


図1 適用領域の設定方法（k近傍法）