

# ラマン分光スペクトルと機械学習および説明可能な人工知能を用いた異常ヘモグロビンの自動検出法

## Automated Detection of Abnormal Hemoglobins from Raman Spectra using Machine Learning and Explainable AI

菱沼 義寛<sup>\*1,\*2</sup>, 孫 宜蒙<sup>\*1</sup>, 今井 正治<sup>\*1</sup>

<sup>(\*1)</sup> 京都情報大学院大学, <sup>(\*2)</sup> 株式会社 オーイーエムシステム

Yoshihiro Hishinuma<sup>\*1,\*2</sup>, Yimeng Sun<sup>\*1</sup> and Masaharu Imai<sup>\*1</sup>

<sup>(\*1)</sup> The Kyoto College of Graduate Studies for Informatics, <sup>(\*2)</sup> OEM Systems Co., Ltd.

### Abstract

This paper proposes a Raman spectroscopy-based framework for classifying abnormal hemoglobin with reduced device dependence and improved chemical interpretability. The authors trained multiple machine-learning classifiers on 1,200 spectra from purified HbA0, HbS, and HbC samples, then used SHAP to identify the wavenumber regions most influential to predictions. These SHAP-derived regions were mapped to established Raman peaks associated with hemoglobin's heme group and amino acid residues, incorporating domain knowledge into model interpretation. The framework matched previously reported performance and achieved higher accuracy across classifiers. Overall, the results suggest that combining Raman spectroscopy with explainable AI can enable robust, reproducible, and interpretable abnormal hemoglobin detection for future clinical screening and diagnostic workflows.

Keywords: Raman Spectral Analysis, Machine Learning, Explainable AI (XAI), SHAP, Principal Component Analysis, Abnormal Hemoglobin Classification

## 1. はじめに

近年、人工知能の発展によりラマン分光データの解析手法が高度化し、応用研究が活発化している。

従来は専門的な知識や多くの時間を要したスペクトル解析が自動化され、より高精度かつ迅速に行えるようになってきた。そのなかでも特に医療・材料科学分野においては応用研究が広がっており、病原体の迅速識別や材料特性評価の自動化など、実用化を見据えた取り組みが進んでいる。医療検査の分野においても、ラマン分光を用いた生体分子の非侵襲分析が注目されており、診断の効率化に寄与する可能性が示唆されている。

一方、糖尿病診断の指標として広く用いられるヘモグロビン A1c (HbA1c) 測定においては、依然として臨床的な課題が残されている。HbA1c は測定法によって異常ヘモグロビンの影響を受けることがあり、偽低値や偽高値を示す例が報告されている

[1, 2]。そのため、正確な診断には異常ヘモグロビンの存在を迅速かつ正確に検出することが不可欠である。本研究では、ラマン分光による異常ヘモグロビン検出に着目し、分類の頑健性ととも、AI が示す分類根拠の化学的妥当性を検証可能な解析フローを構築する。期待される成果は、ヘモグロビンの構造情報に基づき、分類に寄与した特徴量を化学的に説明できる点にある。本研究の貢献は、(1) 異なる装置条件下で取得したデータセットでも安定して機能する装置非依存的なラマン分光分類モデルを構築した点、(2) SHAP を用いて分類根拠を可視化し、その化学的妥当性を検証する枠組みを提示した点にある。

## 2. 先行研究

異常ヘモグロビンは HbA1c 測定に干渉し、偽高値あるいは偽低値を引き起こすことが報告されている。これらの測定誤差は診断精度を損なう可能性が

あり、臨床的に重大な影響を及ぼす場合もある [3]。現行のスクリーニング法としてクロマトグラム波形の精査 [4] や血糖値との相関分析 [5] が用いられるが、いずれも手作業であり検査担当者の負担が大きく、異常ヘモグロビンを迅速かつ自動的に検出する手法が求められている。

異常ヘモグロビンの検査法としては、電気泳動 [6]、高速液体クロマトグラフィ [7]、質量分析 [8]、遺伝子解析 [9] など多様な手法が提案されているが、コストが高い、操作が煩雑などの制約がある。

そこで、ラマン分光は分子固有の振動情報を非破壊的に取得できることから、医療分野での応用が進んでいる。Abbasi ら [10] は液状のヘモグロビン精製品をサンプルとしてラマンスペクトルを取得し、機械学習により各ヘモグロビンを分類する手法を報告した。しかし、これらの研究は単一装置・単一条件での評価に限られており、装置依存性の影響は十分に検証されておらず、分類モデルの判断根拠についても検討が十分であるとは言えない。

## 3. 方法

### 3.1. 材料およびデータ取得

本研究で使用した材料およびラマンスペクトルデータの取得方法について述べる。

#### 3.1.1. 材料

深凍結保存された HbA0, HbS, HbC を特定非営利活動法人病態解析研究所より入手した。これらの試料は、ヘモグロビン濃度 0.7 mg/mL に調製した。各サンプルは測定直前に、0.45  $\mu\text{m}$  サンプル前処理用フィルター (GL CHROMATODISK 13S, ジーエルサイエンス株式会社) でろ過し、不溶性成分を除去した。

#### 3.1.2. ラマンスペクトル測定

サンプルは底面が平坦なウェルプレートに 100  $\mu\text{L}$  ずつ分注した。ラマンスペクトルの取得にはレーザーラマン顕微鏡 RAMANtouch (ナノフoton社製) を使用した。対物レンズは 10  $\times$  を用い、励起光には 532 nm を選択した。出力約 0.4 W の 532 nm レーザービームをライン照明モードで試料に照射した。分光には 1200 本/mm の回折格子を用い、測定スペクトル範囲は 600–1700 $\text{cm}^{-1}$  とした。測定に際しては、レーザー光が最適に集光されるよう X・Y・Z 各軸方向で焦点調整を行い、各測定で 400 点のスペクトルを取得した。すべての測定は室温で

行った。

#### 3.1.3. データ収集

各クラスは純粋なヘモグロビンサンプルから取得したラマンスペクトルで構成し、サンプル当たり 400 スペクトル、合計 1,200 スペクトルのデータセットを作成した。データセットは、A 列に波数、B 列以降に各ラマンスペクトルを格納した CSV 形式で保存した。なお、すべてのデータは同一の測定条件下で、同一日に取得した。

### 3.2. データ前処理および特徴量処理

本研究では、宇宙線スパイク、蛍光バックグラウンド、ならびに強度スケールのばらつきに対して、Python を用いた自作プログラムにより補正した。

#### 3.2.1. 宇宙線スパイクの除去

宇宙線スパイクの除去には、自己相関解析とウェーブレット変換に基づく手法 [11] を用いた。この手法では、まずスペクトルの自己相関関数を計算し、周囲のラマンピークと比較して極端に急峻で孤立した成分を持つ領域を検出する。宇宙線スパイクは、レーザーラマン顕微鏡の CCD 検出器に入射した高エネルギー粒子が引き起こすランダムな鋭いピークであり、通常のラマン散乱とは異なる時間的・空間的特性を示すため、自己相関解析により識別する。続いて、検出されたスパイク領域に対してウェーブレット変換を適用し、急峻な成分のみを選択的に除去することで、周囲のスペクトル形状を保持したまま補正を行う。ウェーブレット変換は、信号を周波数成分ごとに分解し、スパイクに特有の高周波成分を分離できるため、ラマンピークを損なうことなくスパイクのみを除去する。

#### 3.2.2. 蛍光バックグラウンドの補正

ラマン分光では、サンプル中の蛍光物質が励起光により発光し、スペクトル全体に緩やかに上昇する成分として重なって現れることがある。この成分は蛍光バックグラウンドと呼ばれ、ラマン散乱よりも強度が大きい場合が多く、適切に補正しないとラマンピークが埋もれ、解析精度を著しく低下させる。本研究では、蛍光バックグラウンドの補正に 7 次多項式によるベースライン推定 [12] を用いた。この手法では、スペクトル全体の緩やかな変動を多項式で近似し、蛍光成分を推定する。推定されたベースラインを元のスペクトルから差し引くことで、蛍光成分を除去し、ラマン散乱に由来するピークを強調する。

### 3.2.3. 正規化

スペクトル間の強度スケールのばらつきの補正には、以下の3種類の正規化を用いた。

#### (1) Min-Max 正規化

スペクトルの最小値を0, 最大値を1に線形変換する方法 [14] である。スペクトル  $\mathbf{x}=(x_1, x_2, \dots, x_n)$  の各成分  $x_i$  を最小値  $x_{\min}$  と最大値  $x_{\max}$  を用いて次式で正規化する。

$$x_{i,norm} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

スペクトルの強度範囲を統一し、全波数点を同一レンジに揃えることで、サンプル間の比較を安定して行うことができる。

#### (2) Vector 正規化

スペクトル全体をベクトルとして扱い、その長さを1にスケールする方法 [13] である。スペクトル  $\mathbf{x}$  をそのユークリッドノルム  $\|\mathbf{x}\|$  で除することで正規化スペクトル  $\mathbf{x}_{norm}$  を得る。

$$\mathbf{x}_{norm} = \frac{\mathbf{x}}{\|\mathbf{x}\|}, \quad \|\mathbf{x}\| = \sqrt{\sum_{i=1}^n x_i^2}$$

スペクトル形状と相対的な変動量を保持したまま強度差を補正し、全体的なスケール変動を均一化することで、ピーク比や特徴量間の関係性に基づく比較を安定して行うことができる。

#### (3) Z-score 正規化

各波数点の値から平均を引き、標準偏差で割ることで、スペクトルを平均0, 標準偏差1のスケールに変換する方法 [15] である。スペクトル  $\mathbf{x}=(x_1, x_2, \dots, x_n)$  の各成分  $x_i$  を、平均  $\mu$  と標準偏差  $\sigma$  を用いて次式で正規化する。

$$x_{i,norm} = \frac{x_i - \mu}{\sigma}$$

測定条件による強度のばらつきを抑え、波数点ごとの相対的な変動を強調することで、特徴量間の関係性に基づく比較を安定して行うことができる。

### 3.2.4. 主成分分析による次元削減

ラマンスペクトルは数百から数千の波数点を特徴量として持つ高次元データである。分類性能の向上と過学習の抑制を目的として、主成分分析 (Principal Component Analysis, PCA) を用いて次元削減を行った。PCA は線形次元削減手法であり、データの分散を最大化する直交基底を抽出することで、主要な情報を保持したまま低次元空間へ射影することができる。本研究では、正規化後のスペ

クトル行列に対して PCA を適用し、累積寄与率が十分に高くなる主成分を分類モデルへの入力特徴量として用いた。これにより、ノイズ成分や冗長な特徴量を除去しつつ、スペクトルの主要な変動要因を効率的に抽出した。

### 3.3. 装置依存性に対する頑健性評価

前処理および次元削減を施したスペクトルデータに対して、機械学習モデルを用いた分類を行った。本研究では、Cortes & Vapnik[16] による SVM, Breiman[17] による Random Forest (RF), Chen & Guestrin[18] による XGBoost を用いた。

これらのモデルは、ラマンスペクトルのような多次元かつ非線形性を含むデータに適しており、分類性能と解釈性の両面から有用であると考えられる。本研究では、これらのモデルを用いて装置依存性に対する頑健性を評価した。

各モデルの性能を最大限に引き出し、かつ客観的な比較を行うため、遺伝的アルゴリズム (Genetic Algorithm, GA) を用いたハイパーパラメータの最適化を行った。GA は、生物の進化過程を模した探索手法であり、多数の候補の中から最適解を効率的に導出できる。本研究では以下の手順で最適化を行った。

#### (1) 探索プロセス

初期集団サイズ 50, 最大世代数 50 とし、各個体は分類器のハイパーパラメータを遺伝子として保持した。

#### (2) 適応度評価

各個体の適応度には、層化交差検証 (Stratified K-Fold, K=3) による平均分類精度を用いた。

#### (3) 進化戦略

トーナメント選択および一様交叉を採用し、突然変異率は 0.2 とした。また、10 世代連続で適応度が改善しない場合には探索を終了した。

本研究では、3種類の正規化手法 (Min-Max, Vector, Z-score) のそれぞれに対して独立に GA による最適化を実施した。

### 3.4. SHAP による分類根拠の妥当性検証

分類モデルがどの波数領域を根拠として判断しているかを明らかにするため、特徴量重要度および Lundberg & Lee[19] によって提案された SHAP (SHapley Additive exPlanations) 値を用いて分類根拠の可視化を行った。

SHAP 値は、各特徴量（波数）が分類結果に与える寄与度を定量化できる手法であり、モデルの解釈性を高める上で有用である。得られた SHAP 値をラマンスペクトル上にマッピングすることで、分類器が参照している波数領域がヘモグロビンの既知の化学構造や振動モードと整合するかを評価した。

この解析により、分類性能だけでなく、分類根拠の化学的妥当性を併せて検証し、信頼性の高い解析フローの構築を目指した。

## 4. 結果

### 4.1. スペクトル取得と前処理の効果

本研究で取得した HbA0, HbS, HbC のラマンスペクトルには、共通して宇宙線に起因する鋭利なスパイクノイズおよび強い蛍光バックグラウンドが確認された。図 1(a) に前処理前のスペクトルの一例を示す。これら前処理前のデータでは、微弱なラマン散乱光がバックグラウンドノイズに埋もれており、ヘモグロビン分子の構造に由来する固有のピークを判別することは困難であった。

これに対し、3.2.1 項および 3.2.2 項で述べた宇宙線除去と 7 次多項式によるベースライン補正を施したところ、補正後のスペクトルは図 1(b), (c)

に示すように改善された。補正後のスペクトルでは、ヘモグロビンのポルフィリン環の振動モードに対応する  $1370\text{cm}^{-1}$  付近の  $\nu_4$  バンドに加え、 $1500\text{--}1650\text{cm}^{-1}$  に位置する骨格伸縮振動 ( $\nu_{11}$ ,  $\nu_{10}$ ,  $\nu_2$ ) が明瞭に確認された。

### 4.2. 正規化手法による強度変動の抑制

前節で得られた補正後のラマンスペクトルは、レーザー焦点のズレや試料濃度の不均一性に起因して、絶対強度にサンプル間のばらつきが生じていた。これに対し、3.2.3 項で述べた正規化処理を適用した結果、図 2 に示すように、強度スケールが統一された。Min-Max 正規化（図 2 上段）では、各スペクトルの強度が 0 から 1 の範囲にスケールされた。Vector 正規化（図 2 中段）では、スペクトル全体のユークリッドノルムを 1 に揃えられた。また、Z-score 正規化（図 2 下段）では、各波数点の平均を 0、標準偏差を 1 となるよう標準化された。

### 4.3. 主成分分析を用いた特徴抽出

正規化後のラマンスペクトルに対し、主要特徴量を抽出するため PCA を適用した。3.2.4 項の基準に従い、累積寄与率が 85% に達するまでの主成分を各モデルの入力特徴量として採用した。結果、累積寄与率を満たす主成分数は Min-Max 正規化で

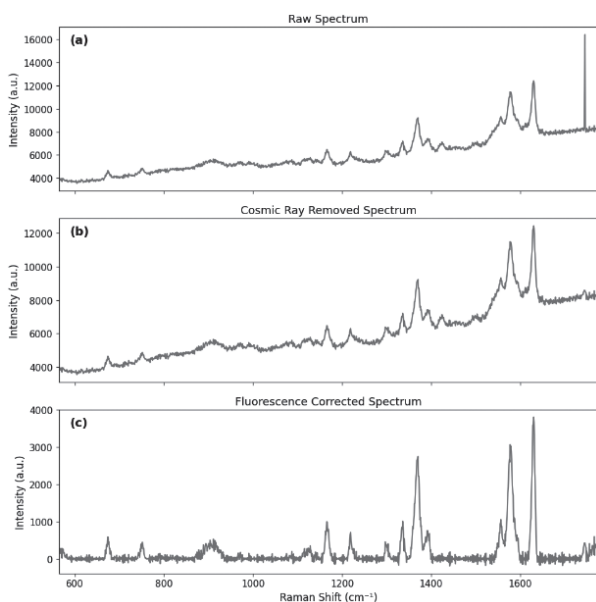


図 1 前処理前後のラマンスペクトルの比較  
(a) 生スペクトル, (b) 宇宙線除去後, (c) 蛍光補正後  
Figure 1. Comparison of Raman spectra before and after preprocessing.

(a) Raw spectrum, (b) after cosmic-ray removal, (c) after fluorescence correction.

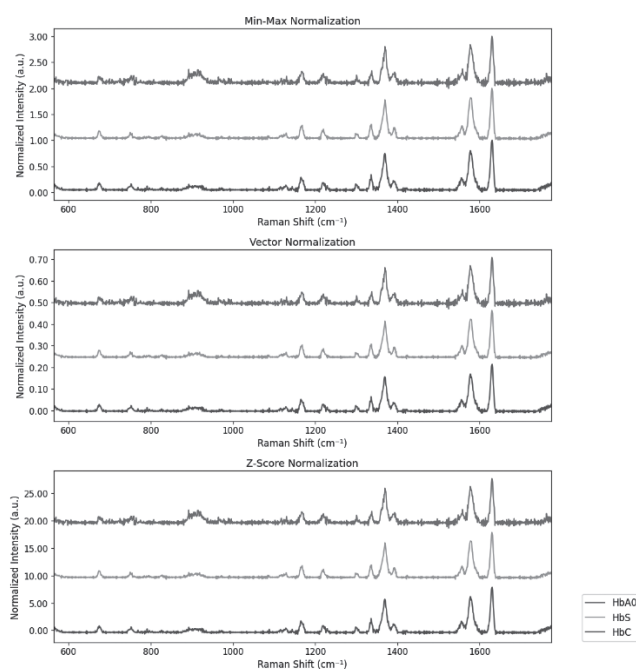


図 2 純粋な HbA0, HbS, HbC の正規化ラマンスペクトル  
Figure 2. Normalized Raman spectra of purified HbA0, HbS, and HbC.

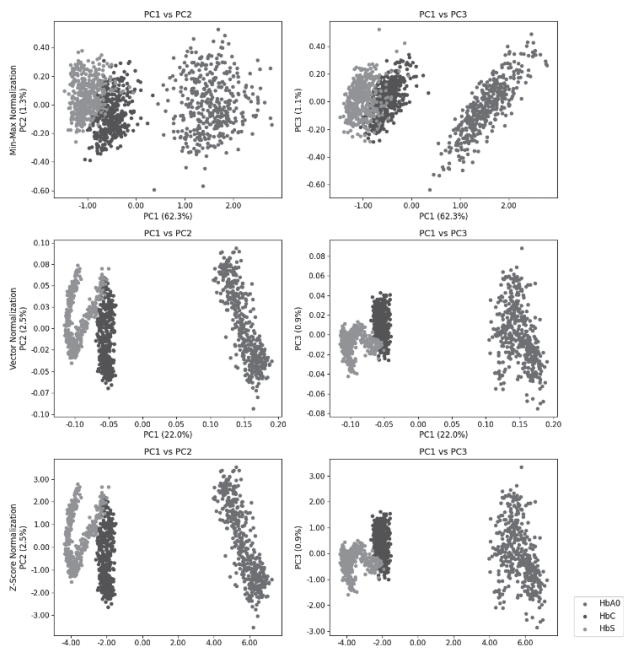


図3 PCAスコアプロット  
Figure 3. PCA score plot.

163, Vector 正規化および Z-score 正規化で 308 となり, 正規化手法によってデータの分散構造に差異が生じることが示された。図 3 に各正規化データにおける主成分スコアの分布を示す。Min-Max 正規化 (図 3(a)) では, PC1 の寄与率が 62.3% と高く, データの主要な分散が第一主成分に集約していた。一方で, Vector および Z-score 正規化 (図 3(b), (c)) では PC1 の寄与率が約 22% に留まり, 情報は複数の主成分に分散する傾向があった。

スコア分布の可視化結果より, いずれの正規化手法においても正常型の HbA0 は変異型の HbS および HbC と明確に分離されることが確認された。一方で, HbS と HbC の分布には一部で重なりが認められた。

表 1 正規化手法ごとの最適化モデルにおける分類性能

Table 1. Classification performance of the optimized model for each normalization method.

Model	Normalization	Accuracy	Precision	Recall	F1-score	AUC
SVM	Min-Max	0.9958	0.9959	0.9958	0.9958	1.0000
	Vector	1.0000	1.0000	1.0000	1.0000	1.0000
	Z-score	1.0000	1.0000	1.0000	1.0000	1.0000
RF	Min-Max	0.9875	0.9880	0.9875	0.9875	0.9996
	Vector	0.9792	0.9804	0.9792	0.9791	0.9998
	Z-score	0.9792	0.9804	0.9792	0.9791	0.9994
XGBoost	Min-Max	0.9917	0.9919	0.9917	0.9917	1.0000
	Vector	0.9833	0.9841	0.9833	0.9833	0.9999
	Z-score	0.9833	0.9841	0.9833	0.9833	0.9998

#### 4.4. 装置依存性に対する頑健性評価

PCA によって抽出された主成分特徴量を用い, SVM, RF, XGBoost の 3 種類のモデルによる HbA0, HbS および HbC の 3 群を対象とした 3 クラス分類を行った。各モデルの学習に際して, 3.3.5 項で述べた GA を用い, 正規化手法ごとにハイパーパラメータの最適化を行った。表 1 に, 最適化された各モデルの分類性能を示す。すべてのモデルにおいて, Accuracy 0.975 以上という極めて高い識別精度が得られた。特に SVM では, Vector 正規化および Z-score 正規化を適用した場合に Accuracy, Precision, Recall, F1 スコア, AUC のすべてにおいて 1.0000 を達成した。

#### 4.5. SHAP による分類根拠の妥当性検証

構築した分類モデルがどの特徴量を根拠として判断しているか可視化するため, SHAP を用いた寄与度解析を行った。本解析では, 3.2.4 項で述べた PCA による次元削減プロセスを考慮し, 得られた主成分ごとの SHAP 値を主成分負荷量と統合して波数空間へ再投影することで, モデルの予測に対する各波数の貢献度を算出した。表 2 に, 各モデルにおいて高い寄与を示した主要な波数を示す。すべてのモデルおよびすべての正規化手法において共通して  $910 \text{ cm}^{-1}$  が極めて高い寄与を示した。また, RF および XGBoost では, ポルフィリン環の酸化状態や電子密度を反映する  $1580 \text{ cm}^{-1}$ , およびスピン状態・配位状態に敏感な  $1627 \text{ cm}^{-1}$  についても高い寄与を示した。さらに, RF では Min-Max 正規化において  $1218 \text{ cm}^{-1}$  が重要特徴量として抽出された。

表 2 SHAP 解析により特定された各モデルの重要波数領域

Table 2. Important wavenumber regions identified for each model by SHAP analysis.

モデル	特徴量 (波数)	適用正規化
SVM	$910 \text{ cm}^{-1}$	Min-Max / Vector / Z-score
RF	$910 \text{ cm}^{-1}$	Min-Max / Vector / Z-score
RF	$1218 \text{ cm}^{-1}$	Min-Max
RF	$1580 \text{ cm}^{-1}$	Min-Max / Z-score
RF	$1627 \text{ cm}^{-1}$	Min-Max / Vector / Z-score
XGBoost	$910 \text{ cm}^{-1}$	Min-Max / Vector / Z-score
XGBoost	$1580 \text{ cm}^{-1}$	Min-Max
XGBoost	$1627 \text{ cm}^{-1}$	Min-Max

## 5. 考察

本研究では、ラマン分光法と Explainable AI (XAI) を組み合わせ、装置や測定環境の差異に左右されず、かつ化学的妥当性を担保した異常ヘモグロビン解析基盤を確立した。

表1に示すように、独自データセットに対してGAを用いた最適化フローを適用した結果、すべてのモデルにおいて97%を超える高い分類精度が得られた。先行研究とは異なる装置および測定条件下で取得したスペクトルを用いても高精度が維持されたことは、構築したモデルが装置固有のバイアスではなく、ヘモグロビン分子に基づく分光学的特徴を抽出できていることを示唆している。本データセットは、1サンプルあたり400回の測定により生じる統計的なばらつきを含んでおり、装置間の個体差、測定条件の差異、日内変動、およびランダムノイズを内包している。この多様な変動要因が存在する条件下で高精度を維持できたことは、適切な前処理によってノイズが抑制され、相対的なピーク強度比や分子振動パターンといった特徴量が、本手法の頑健性を裏付けている。さらに、Vector正規化やZ-score正規化の適用により、ベースラインの変動や強度の絶対値に依存せず、スペクトル形状の微細な差異を強調または保持できたことが、装置に依存しない特徴抽出を可能にしたと考えられる。

表2に示すように、分類モデルごとに重要視する波数域が異なるという知見が得られた。SVMが $920\text{cm}^{-1}$ 付近の特定の領域に強く依存するのに対し、RFやXGBoostはポルフィリン環由来の $\text{C}=\text{C}$ 伸縮振動に対応する $1580\text{cm}^{-1}$ や $1627\text{cm}^{-1}$ といった広範囲の波数域を参照した。これらの波数はヘムの電子状態や配位状態に由来する既知の指標であり[20]、RFやXGBoostがこういった変化を捉えて分類を行っていることを示唆している。RFは多数の決定木から構成されるアンサンブル学習であり、非線形性や特徴量間の相互作用を自然に表現できるモデルである。そのため、異常ヘモグロビン特有のポルフィリン環の歪みやヘムの電子状態の変化といった、複数の波域にまたがる変化を検知できたと考えられる。SHAPを通じた化学的根拠の付与は、将来的な臨床応用に向け重要であるといえる。

一方で、3種類のモデルで共通して重要視された $920\text{cm}^{-1}$ 付近の領域は、タンパク質の骨格振動に由来する可能性はあるが、ヘモグロビンに特異的な

指標としての既知の報告は確認されなかった。この領域がヘモグロビンの立体構造に関連する指標となるのか、あるいはサンプル調製、測定、前処理のプロセスで生じた人工的な歪みであるのかについては、今後の検証が必要である。また、全血サンプルのような生物試料は、脂質や代謝物等が混在する複雑なスペクトルを呈するため、精製サンプルを用いた本モデルの知見をそのまま適用することには限界がある。今後は、MCR-ALS (Multivariate Curve Resolution-Alternating Least Squares) [21] 等を用いて複合スペクトルを各成分へと分解し、本研究で得られた純物質スペクトルの知見を適用することが有効と考えられる。

## 6. まとめと今後の課題

本研究では、精製した3種類のヘモグロビンを対象として独自に取得したラマンスペクトルデータセットを用い、機械学習モデルに対してSHAPによる特徴量の可視化を行い、分類に寄与した特徴量とヘモグロビンの分子構造に関する知見との整合性を評価した。その結果、モデルごとに参照する波数域が異なることが判明した。特にRFやXGBoostでは異常ヘモグロビンに特有のヘム周辺の構造変化を反映した波数が判断に大きく寄与しており、化学的な妥当性が示された。

今後は、全血サンプルのような複雑なスペクトルへの対応を見据え、主観的判断に依存しない深層学習モデルの適用など、より実用的な検出手法の構築を検討していきたい。

### 謝辞

本研究の遂行にあたり、貴重なご助言をいただいた京都情報大学院大学の中口孝雄教授および上野未貴講師に深く感謝します。また、本研究の遂行にあたり、学業と業務の両立にご理解を賜り、研究活動に温かいご支援をいただきました株式会社オーイーエムシステム代表取締役社長 白石貴氏に深謝します。

### 参考文献

- [1] Mingyang Li, Song Ge, Xin Shu Xiongjun Wu, Haiyan Liu, Anping Xu, Ling Ji., "Interference of hemoglobin variants with HbA1c measurements by six commonly used HbA1c methods", *Laboratory Medicine*, 55, 6, pp. 708–712, 2024.
- [2] Khajuria A, Scheibe B, "Effect of Elevated levels of

Hemoglobin F on HbA1c Measurements: Evaluation of Three HbA1c Assays”, *Acta Medica International* 2 (2), 85-88, 2015.

- [3] 清水彩洋子, 平良暁子, 畑崎聖弘, 馬屋原豊, 平良真人, 古賀正史, “HbA1c が偽性高値を示したために経口血糖降下薬の投与を受けた非糖尿病異常ヘモグロビンの2例”, *糖尿病*, 58, 2, pp. 121-127, 2015.
- [4] 木下真紀, “意外に多いヘモグロビン異常症 –ヘモグロビンA1c値に与える影響–”, *天理医学紀要*, 2, pp. 88-95, 2020.
- [5] 佐藤智子, 白幡季大, 木下正行, 小森崇司, 三代久美子, 中道陽子, 木下由美子, 矢口通子, 栗原達哲, 眞崎正, 山縣文夫, 河津捷二, 及川孝光, “当施設受診者における異常ヘモグロビン 検出状況について”, *総合健診*, 46, 5, pp. 468-479, 2019.
- [6] Nathalie Mario, Bruno Baudin, Arnaud Bruneel, Jacques Janssens, Michel Vaubourdolle. “Capillary zone electrophoresis for the diagnosis of congenital hemoglobinopathies”, *Clinical Chemistry*, 45, 2, pp. 285-288, 1999.
- [7] Jean Riou, Christian Godart, Didier Hurtrel, Mireille Mathis, Catherine Bimet, Josiane Bardakdjian-Michau, Claude Pre’hu, Henri Wajcman, Fre’de’ric Galacte’ros, “Cation-exchange HPLC evaluated for presumptive identification of hemoglobin variants”, *Clinical Chemistry*, 43, 1, pp. 34-39, 1997
- [8] Roger Théberge, Sergei Dikler, Christian Heckendorf, David H. K. Chui, Catherine E. Costello, Mark E. McComb., “MALDI-MS Mass Spectrometry Analysis of Hemoglobin Variants: a Top-Down Approach to the Characterization of Hemoglobinopathies”, *J. Am. Soc. Mass Spectrom*, 26, pp. 1299-1310, 2015.
- [9] Bonaventura J, Riggs A, “Hemoglobin Kansas, a human hemoglobin with a neutral amino acid substitution and an abnormal oxygen equilibrium”, *J Biol Chem*, 243, 5, pp. 980-991, 1968.
- [10] Sara Abbasi, Mehdi Feizpour, Ilse Weets, Qing Liu, Hugo Thienpont, Francesco Ferranti, Heidi Ottevaere, “Classification of hemoglobin fractions in the liquid state using Raman spectroscopy combined with machine learning”, *Microchemical Journal*, 194, 109305, 2023.
- [11] Augusto Maury, Reynier I. Revilla, “Autocorrelation Analysis Combined with a Wavelet Transform Method to Detect and Remove Cosmic Rays in a Single Raman Spectrum”, *Appl Spectrosc*, 69, 8, pp. 984-992, 2015.
- [12] Chad G. Atkins, Kevin Buckley, Michael W. Blades, Robin F.B. Turner, “Automated Method for Subtraction of Fluorescence from Biological Raman Spectra”, *APPLIED SPECTROSCOPY*, 57, 11, pp. 1363-1367, 2003.
- [13] Rasmus Broa, Age K. Smilde, “Principal component analysis”, *Anal. Methods*, 6, 2812, 2014.
- [14] Wandan Zeng, Qi Wang, Zhiping Xia, Zhiping Li, Han Qu., “Application of XGBoost Algorithm in The Detection of SARS-CoV-2 “Using Raman Spectroscopy”, *Journal of Physics: Conference Series*, 1775, 012007, 2021.
- [15] Timothy W. Randolph, “Scale-based normalization of spectral data”, *Cancer Biomarkers*, 2, 135-144, 2006.
- [16] C. Cortes, V. Vapnik, “Support-vector networks”, *Machine Learning*, 20, 273-297, 1995.
- [17] L. Breiman, “Random Forests”, *Machine Learning*, 45, 1, pp. 5-32, 2001.
- [18] T. Chen, C. Guestrin, “XGBoost: A Scalable Tree Boosting System”, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.785-794, 2016.
- [19] Lundberg, S. M., Lee, S.-I., “A Unified Approach to Interpreting Model Predictions”, *arXiv:1705.07874*, 2017.
- [20] B. BENKO, “Resonance Raman studies of nitric oxide binding to ferric and ferrous hemoproteins: Detection of Fe(III)-NO stretching, Fe(III)-N-O bending, and Fe(II)-N-O bending vibrations”, *Proc. Natl. Acad. Sci USA*, 80, pp. 7042-7046, 1983.
- [21] Keita Iwasaki, “Raman spectroscopic studies coupled with MCR-ALS applied on some biomedical systems”, *Doctoral dissertation, The United Graduate School of Agricultural Sciences, Tottori University*, 2021.

◆著者紹介

**菱沼 義寛 Yoshihiro Hishinuma**

京都情報大学院大学 大学院生  
株式会社オーイーエムシステム

**孫 宜蒙 Yimeng Sun**

京都情報大学院大学 講師

**今井 正治 Masaharu Imai**

京都情報大学院大学 教授