4. 結果

カラフルな絵

中程度の精度で自動的に生成された説明

図　Fpocket[4]から得られた特徴量の分布(青: クリプトサイトになり得る凹み、赤: その他の凹み)の比較。

表　各特徴量とコルモゴロフ・スミルノフの2標本統計に基づくP値の関係

|  |  |
| --- | --- |
| 特徴量名 | P値 |
| Hydrophobicity score | 0.004505535750710443 |
| Mean alp. sph. solvent access | 0.012702623743886199 |
| Number of Alpha Spheres | 0.15383457971242187 |
| Score | 0.16894901636395654 |
| Mean local hydrophobic density | 0.16995554937081192 |
| Polarity score | 0.24012107232767865 |
| Alpha sphere density | 0.24012107232767865 |
| Proportion of polar atoms | 0.3787916710900978 |
| Mean alpha sphere radius | 0.39169455136082354 |
| Cent. of mass - Alpha Sphere max dist | 0.4074716655962288 |
| Polar SASA | 0.628495301602217 |
| Volume | 0.628495301602217 |
| Total SASA | 0.6297862188632384 |
| Apolar SASA | 0.7489562619763734 |
| Apolar alpha sphere proportion | 0.7859979604370235 |
| Druggability Score | 0.8179280488163592 |
| Flexibility | 0.848614167291587 |
| Volume score | 0.9179177566857942 |
| Charge score | 0.999999561098095 |

クラスタリング

グラフ, 散布図

自動的に生成された説明グラフ, 散布図

自動的に生成された説明

グラフ, 折れ線グラフ

自動的に生成された説明屋内, テキスト, 地図, テーブル が含まれている画像

自動的に生成された説明

図　データセットをPCAし、PC1とPC2について可視化(青：クリプトサイトになり得る凹み、赤：その他の凹み)(左上)、データセットをPCAし、PC1とPC2について可視化(青：学習データ、赤：テストデータ)(右上)、累積寄与率の可視化(左下)、PC1とPC2における各特徴量の寄与度の可視化(右下)

機械学習モデルの予測

構成したデータセットを学習データ175個、テストデータを36個に分けた。機械学習モデルとしてRandom Forest、XGBoost、LightGBM、SVMを用いて学習し、性能を比較した。その結果、SVMが一番性能がよく、テストデータについてF1\_score: 71.0%の性能を達成した。

グラフ

自動的に生成された説明

図　混合行列（SVM）

機械学習モデルが正答した場合(図２)と誤答した場合(図３)のタンパク質のポケット周辺の表面構造を確認した。現状としてモデルが予測を誤答する場合はアポ構造においてクリプトサイトになりうる凹みが浅く、その他の凹みと判定を誤ったと考えられる。

リンゴの絵

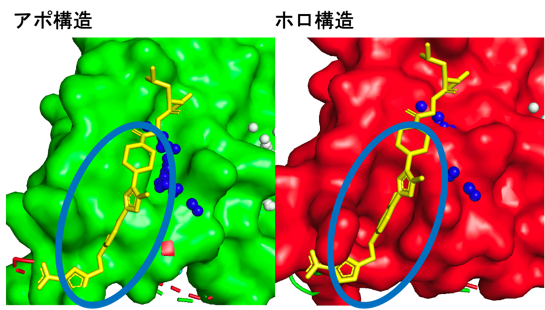
中程度の精度で自動的に生成された説明　　

図３：モデルが予測を誤答(その他の凹みと予測)したタンパク質のポケットの例（アポ構造: 1Z92A、ホロ構造: 1PY2A、リガンド: FRH）

図２：モデルが予測を正答したタンパク質のポケットの例（アポ構造: 1BSQA、ホロ構造: 1GX8A、リガンド: RTL）

機械学習モデルの因子分析

また各モデルが学習において重要と判断した特徴量を可視化した(図４)。その結果、各モデル共にHydrophobicity score、Alpha sphere density、Polarity scoreが重要特徴量の上位であることがわかった。

グラフ, じょうごグラフ

自動的に生成された説明　グラフ

自動的に生成された説明

図４：SVMの重要特徴量の可視化

ダイアグラム

自動的に生成された説明ダイアグラム

自動的に生成された説明

ダイアグラム

中程度の精度で自動的に生成された説明