

個体の生体応答記述を志向した毒性病理画像の潜在表現に関する基礎研究

○ 水野 忠快^{1,2}, 前寺正太郎¹, 森田勝久¹, 楠原洋之¹

¹東京大学 大学院薬学系研究科 分子薬物動態学教室, ²統計数理研究所 統計思考院

Mizuno-group

tadahaya@gmail.com

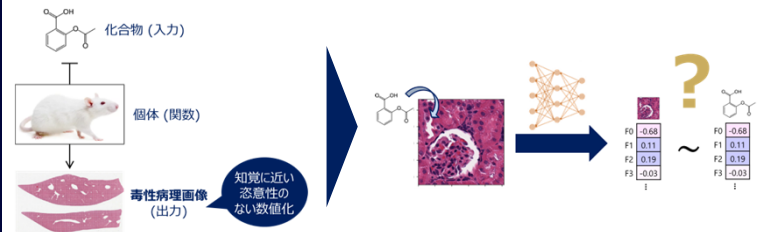


github organization



■ ①背景

毒性病理画像は個体への生体応答を記述する貴重なデータ



- ✓ 病理画像は個体の状態を記述する重要な情報であり、疾患診断や医薬品開発における安全性評価に不可欠。
- ✓ 画像は人間の知覚に近い情報を提供するため、恣意性なく化合物の作用を数値化可能。
- ✓ 特に毒性病理画像(化合物が投与された個体の病理組織画像)は、個体が化合物に対して示す生体応答を反映。
- ✓ 化合物と個体を結びつけるデータ駆動型解析研究の対象として極めて重要なデータと言える。

■ ③方法

使用データ

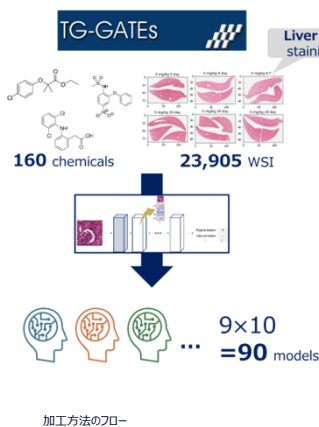
- ✓ Open TG-GATES, 肝病理画像: 23,905WSI
- ✓ 化合物数: 160
- ✓ 病理所見数: 44

加工方法

- ✓ 512x512 pixelsのパッチ画像へとWSIを分割して解析に使用。
- ✓ まず44所見各々についてWSIへのラベルに基づき弱教師あり学習にてパッチ単位の所見有無の推定を実施。
- ✓ 同結果に基づき、所見有のスコアが0.5を超える8所見に絞る、以下の解析を実施。

性能評価方法

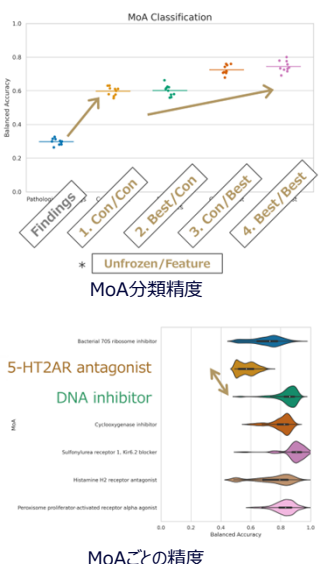
- ✓ ImageNetで事前学習済みのEfficientNetB4を使用。
- ✓ ①どのレイヤーから特徴量を抽出するか、②どのレイヤーまで事前重みを利用するかを振り、 $9 \times 10 = 90$ modelsを構築。
- ✓ 所見についてLOOを実施し、7所見を補助タスクとして学習した後外していた8所見目を推定する性能を評価。
- ✓ 画像数の多い7つのMoAに関する分類性能を評価。



■ ⑤結果2

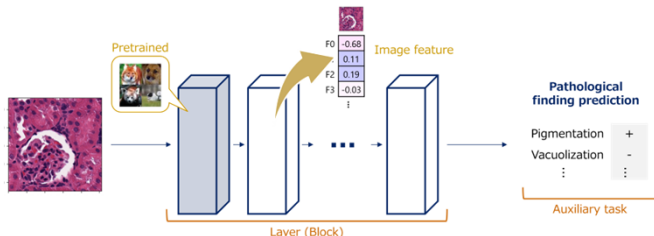
取得した潜在表現による下流タスクの性能評価(MoAの分類)

- ✓ 特徴量抽出層と凍結層、それぞれについて、補助タスクにて最良であったモデル(Best)を用いて下流タスクの性能を評価。
- ✓ Mode of Action (MoA, 化合物の作用機序)分類を実施。
- ✓ MoAとしては、多くの毒性病理画像に共通していた7種を選定して用いた。
- ✓ 対照群として、病理所見自体を日ベクトルとしたもの、及びImageNetの事前学習モデル(Con)を使用して比較。
- ✓ 特徴量抽出層を最適化した場合に精度向上が認められた。
- ✓ どちらの層もBestなモデルについて、臓器特異性の低いDNA阻害剤は高いスコアを示した一方、臓器特異性の高いセロニン受容体アンタゴニストは比較的低い精度を示した。



■ ②目的

有用な潜在表現に向け、(1)特徴量抽出層、(2)学習時事前重みを活用する層(凍結層)の選定の影響を検討する。

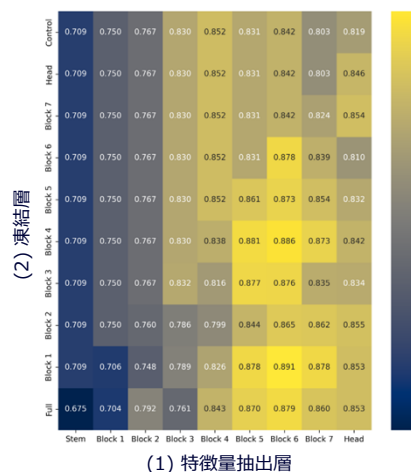


- ✓ 毒性病理画像の潜在表現は、化合物の個体での作用を表すと考えられる。
- ✓ どのようにして有用な潜在表現を取得可能だろうか？
- ✓ 表現学習のうち最もシンプルなアーキテクチャの一つである補助タスクの学習による潜在表現抽出に着目し、以下の問いに答えることを目的とした。
- ✓ (1)特徴量をどの層から抽出するか？
- ✓ (2)ファインチューニング時、事前学習の重みをどの層まで活用するか？ (=ファインチューニング時に勾配を流さない層)

■ ④結果1

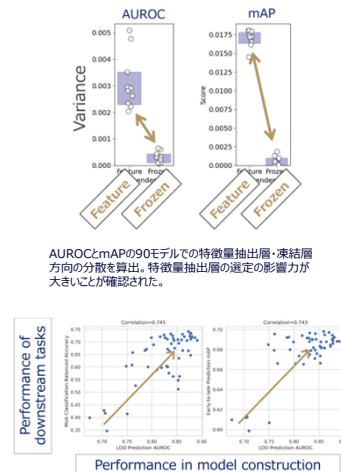
レイヤーの違いが与える補助タスク性能への影響

- ✓ 2つの観点でレイヤーを振って構築した病理所見予測モデルの性能(AUROC)をヒートマップにより可視化。
- ✓ (1)特徴量抽出層としては、出力に近い層を選定した方がスコアが高い傾向にある。
- ✓ (2)凍結層については明確な傾向は認められない。
- ✓ 両者についてはスイートスポットが存在する(特定の軸だけでは決定されない)。
- ✓ 特徴量抽出層と凍結層の組み合わせにより、AUROCは20ポイントの幅がある。



■ ⑥結論

- ✓ 補助タスクベースの毒性病理画像表現学習モデルでは、特徴量抽出層の選定がその精度に大きく影響する。
- ✓ 凍結層はそれほど大きな影響を示さなかったものの、特徴量抽出層との組み合わせではスイートスポットが認められた。
- ✓ 層選定はAUROCにして20ポイント程度影響を与える。
- ✓ 補助タスクの精度は下流タスクの精度と比較的相関する。



潜在表現を得るための補助タスクの性能と、潜在表現の下流タスクの相関を評価。比較的良好な相関が認められた。ただし特にスコアが低いデータの相関が大きいことは否めず、より詳細な検討が必要とされる。