Vol.2015-MPS-103 No.36
Vol.2015-BIO-42 No.36
2015/6/24

SVMとDeep Learningに基づくヒト c-Yes キナーゼ阻害化合物の予測

鈴木 翔吾[1,2,a)]　柳澤 渓甫[1,2,b)]　大上 雅史[1,c)]　石田 貴士[1,2,d)]　秋山 泰[1,2,e)]

概要：新薬開発にはバーチャルスクリーニングという，新薬候補化合物をコンピュータ上で選別する手法が広く用いられている．バーチャルスクリーニングには，標的タンパク質に対する阻害活性が既知の化合物情報を用いるリガンドベースの手法（ligand-based method）があるが，これには様々な機械学習やデータマイニングの手法が用いられている．本研究では，SVM と Deep Learning の 2 種類の学習方法を用いて標的阻害化合物の予測モデルを構築した．また，Deep Learning の出力層に SVM を用いた予測モデルを構築した．そして，これら 3 つの予測モデルをヒト c-Yes キナーゼ阻害化合物の予測に適用し，精度の評価を行った．

キーワード：バーチャルスクリーニング，機械学習，SVM，Deep Learning，c-Yes

Prediction of Human c-Yes Kinase Inhibitors by SVM and Deep Learning

Shogo D. Suzuki[1,2,a)]　Keisuke Yanagisawa[1,2,b)]　Masahito Ohue[1,c)]
Takashi Ishida[1,2,d)]　Yutaka Akiyama[1,2,e)]

Abstract: Virtual Screening (VS) is widely used in the process of a new drug development. A ligand-based method which is widely used in VS uses molecular descriptors of compounds of which inhibition activity for a target protein is proved. In ligand-based methods, many methods of machine learning or data mining are used. In this research, we constructed prediction models for target inhibition compounds by SVM and Deep Learning. In addition to the models, we constructed a prediction model by Deep Learning whose output layer is SVM. Finally, we applied these three models to prediction of human c-Yes kinase inhibitors and evaluated their accuracy.

Keywords: Virtual Screening, Machine Learning, SVM, Deep Learning, c-Yes

1. 導入

新薬開発には有効性や安全性を試験するために非常に長い年月と多大な費用が必要である．このコストを抑えるために，新薬になる可能性が高い化合物を基礎研究の段階でできるだけ多くスクリーニング・生化学実験にかける必要がある．しかし，化合物ライブラリーに含まれる全ての化合物をスクリーニングにかけるのは非常に大きなコストがかかる．そこで，新薬になる可能性の高い化合物をあらかじめコンピュータ上で選定する，バーチャルスクリーニン

[1] 東京工業大学 大学院情報理工学研究科 計算工学専攻，
Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology
[2] 東京工業大学 情報生命博士教育院，
Education Academy of Computational Life Sciences, Tokyo Institute of Technology
[a)] s_suzuki@bi.cs.titech.ac.jp
[b)] yanagisawa@bi.cs.titech.ac.jp
[c)] ohue@bi.cs.titech.ac.jp
[d)] ishida@cs.titech.ac.jp
[e)] akiyama@cs.titech.ac.jp

グという技術が用いられている．

バーチャルスクリーニングには，標的タンパク質に対する阻害活性が既知の化合物情報を用いるリガンドベースの手法（ligand-based method）があるが，これにはサポートベクターマシン（Support Vector Machine: SVM），決定木，ナイーブベイズ，k 最近傍法，ニューラルネットワークといった，様々な機械学習やデータマイニングの手法が用いられている [1]．中でも SVM は他の機械学習アルゴリズムと比較して，計算量は大きいが予測精度は高いという報告がされている [1]．

一方で，音声認識や画像認識といった分野においては深層学習（Deep Learning）が近年広く用いられてきている．音声認識や画像認識のベンチマークテストにおいて，Deep Learning が以前の記録を塗り替えている [2,3]．これらの分野だけでなく，化合物の活性予測においても Deep Learning の有効性が報告されている [4]．標的阻害化合物の予測においても Deep Learning による予測モデルが高い精度を出すことが期待される．

そこで，本研究では SVM と Deep Learning をそれぞれ用いて標的阻害化合物の予測モデルを構築した．また，SVM と Deep Learning を組み合わせた予測モデルは単独で用いた場合よりも高い精度となるのではないかと考え，Deep Learning の出力層に SVM を用いた予測モデルを構築した．そして，これら 3 つの予測モデルをヒト c-Yes キナーゼ阻害化合物の予測に適用し，精度の評価を行った．

ヒト c-Yes キナーゼはチロシンキナーゼの中の Src ファミリーの 1 つであり，ウェストナイル熱を引き起こすウェストナイルウイルスの生活環に影響を与えることが示唆されている [5]．ウェストナイルウイルスに対するワクチンは今のところ存在せず [6]，ワクチン開発のための創薬研究が重要視されている．

先行研究に，構造ベース法（structure-based method）を c-Yes に適用した研究として，C. Ramakrishnan ら（2014）の研究 [7] がある．また，c-Yes キナーゼが属する Src ファミリーに含まれる他のキナーゼを標的タンパク質とし，リガンドベースの手法を用いた研究として J. Witek ら（2013）の研究 [8] がある．この研究では，ナイーブベイズ，SVM，ランダムフォレストといった機械学習アルゴリズムが用いられていた．また，Src ファミリーに含まれるキナーゼ全体を標的タンパク質とし，リガンドベースの手法を用いた研究として M. Zheng（2014）の研究 [9] がある．この研究では，ナイーブベイズが用いられていた．

2. 手法

2.1 サポートベクターマシン

サポートベクターマシン（Support Vector Machine: SVM）は最適分離超平面を求める機械学習の手法である．本研究では Scikit-learn（Ver. 0.15.2）[10] で実装されている SVM を用いた．また，SVM は高次元写像による複雑な決定境界の生成にカーネル関数を用いることで効率的に計算することが可能である．カーネル関数には線形カーネル，多項式カーネル，ガウスカーネルといった様々な関数が存在するが，ガウスカーネルのハイパーパラメータを調整することで，線形カーネルや多項式カーネルを用いたときに得られる決定境界と似たような決定境界を得ることができるため [11]，本研究ではガウスカーネルを SVM のカーネル関数として用いた．

2.2 Deep Learning

Deep Learning は深い層構造のニューラルネットワークを利用する機械学習の手法である．本研究では pylearn2 [12] で実装されている Deep Learning を用いた．

Deep Learning は過学習を起こしやすいという問題があるため，過学習を避けるための様々な研究が行われている．本研究で用いた過学習を避けるためのテクニックを以下に示す．

事前学習

事前学習とは，教師なし学習を用いて Deep Learning における各層の重みの初期値を決定する方法である．事前学習によって決定された重み初期値を用いて fine tuning を行うことで過学習を避けられることが報告されている [13]．事前学習の手法の 1 つとして，Denoising Autoencoder（DA）[14] がある．DA は確率的にノイズが加えられた入力を符号化・復号化した値と，ノイズが加えられる前の入力との誤差を最小化するような重みを求める手法である．本研究では DA を積み重ねた Stacked Denoising Autoencoders（SDA）[15] を事前学習に用いた．SDA における符号化関数には $\tanh(x)$，復号化関数には恒等関数を用いた．また，SDA における誤差関数に正則化項を加えた（スパース正則化）．スパース正則化における誤差関数 $\tilde{E}(\mathbf{w})$ を以下に示す．

$$\tilde{E}(\mathbf{w}) \equiv E(\mathbf{w}) + \beta \sum_{j=1}^{D_y} \mathrm{KL}(\rho||\hat{\rho}_j) \tag{1}$$

ここで，$E(\mathbf{w})$ を元の誤差関数，D_y を中間層のユニット数，$\hat{\rho}_j$ を中間層のユニット j の平均活性度の推定値を表す．ρ と β はスパース正則化におけるハイパーパラメータである．

重みの制約

本研究では，fine tuning における誤差関数に重みの二乗和（ℓ_2 ノルム）で表される正則化項を加えた．また，各層のユニットから確率的に選別されたユニットのみを用いて学習を行うドロップアウト [16] を用いた．ドロップアウトを用いることで，学習時におけるネットワークの自由度を制限し過学習を避けられることが報告されている [16]．

Deep Learning には決定すべきハイパーパラメータが複

数存在するが，特に誤差関数に対する確率的勾配降下法における学習係数は，予測モデルの性能を大きく左右するパラメータであることが報告されている [17]．学習係数を自動的に決定する手法として，AdaGrad [18] および AdaGrad の改良手法である AdaDelta [19] がある．AdaGrad は現在のステップまでに計算された勾配の全てを考慮に入れて学習係数を決定するのに対し，AdaDelta は直近のステップにおける勾配を用いて新しく学習係数を決定するという特徴がある．本研究では AdaDelta を用いて学習係数を自動的に決定した．

2.3 Deep Learning + SVM

本研究では Deep Learning の出力層に SVM を用いた予測モデルの構築も行った．すなわち，Deep Learning によって変換された特徴量が SVM に入力として与えられる予測モデルである．この予測モデルの概念図を**図 1** に示す．

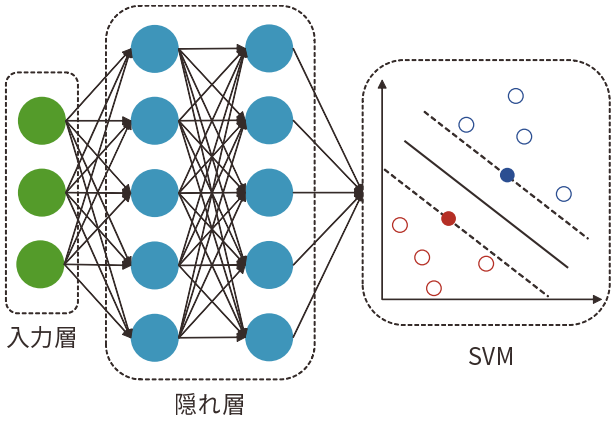

図 1 Deep Learning + SVM の概念図

特徴量変換器として用いる Deep Learning における各層の重みは SDA によって決定した．また，出力層の SVM のカーネル関数にはガウスカーネルを用いた．

3. 評価実験

3.1 データセット

本研究で予測モデルを構築する際に用いた訓練データおよび予測モデルの評価に用いたテストデータの概要を以下に示す．

訓練データ

訓練データには Kinase SARfari [20] から得られた化合物データを用いた．Kinase SARfari は標的タンパク質がキナーゼに特化した化合物が集められている化合物データベースである．本研究では Kinase SARfari から次の 2 つの条件を満たす化合物を取得した．

(1) c-Yes キナーゼを含む，Src ファミリーに含まれるヒトのキナーゼを標的タンパク質としたもの
(2) 活性値が IC_{50} で与えられているもの

IC_{50} は「ある値が定まっている（relation: “=”）」・「ある値より小さい，もしくは以下（relation: “<”，“<=”）」・「ある値より大きい，もしくは以上（relation: “>”，“>=”）」という 3 種類のうちどれか 1 つの報告がされている．ある化合物について同じ標的タンパク質に対する IC_{50} が複数報告されている場合は，それらの平均値を活性値とした．本研究では活性値が厳密に 10 μM より小さくなるときに “Active” のラベルを，厳密に 100 μM より大きくなるときに “Inactive” のラベルを，それ以外のときに “Inconclusive” のラベルを付与した．阻害活性ラベルが “Inconclusive” の化合物については，訓練データから除外した．

テストデータ

テストデータには 2014 年に開催された第 1 回 IPAB コンテスト [21] によって得られたヒト c-Yes キナーゼに対するアッセイ情報を用いた．IPAB コンテストとは，IT 創薬に関する技術を広く浸透させること，IT 創薬への参加者の裾野を広げることを目的とする，創薬プロセスの上流であるヒット化合物の探索をテーマとしたコンテストである．IPAB コンテストのアッセイは Bienta 社が担当し，Promega 社の ADP-Glo kinase assay platform で poly (Glu-Tyr) substrate を使用した Yes kinase スクリーニングのキットが用いられた．以下に具体的なアッセイの方法を示す．

(1) プライマリーアッセイ
全ての化合物を 8 枚の 384 well プレートに 4 well 分ずつ割り振り，10 μM の固定された濃度で，阻害率測定を 4 回（4 well 分）行い，それぞれの化合物について，4 回のアッセイ結果の平均値をとり，以下の基準でバリデーションに進める化合物を選択した．
(a) Bienta 社によりプライマリーヒットと定義された化合物
(b) (a) に該当しない化合物で，阻害率が 30%を越えているもの
(c) (b) に該当しない化合物で，各グループの最大の阻害活性を有する化合物 1 つ

(2) バリデーション
(a)，(b)，(c) に該当する化合物が 1 枚のプレート上で 6 well 分ずつアッセイされた．6 回のアッセイの平均値を用いて最終的な阻害活性の有無が判断された．

本研究では，Bienta 社によりプライマリーヒットと定義された化合物 ((a) に相当する化合物) に “Active” のラベルを付与し，その他の化合物については “Inactive” のラベルを付与した．

特徴量

データセットの各化合物に対し，ケモインフォマティクスのライブラリである RDKit（Ver. 2014/3/1）[22] を用いて分子量や分配係数など 196 種類の化合物情報を計算した．各特徴量は取りうる値の範囲が異なり，取りうる値の範囲が大きい特徴量が予測に対して大きな影響を与えてしまう．この問題を回避するため，z-score を用いて各特徴量のスケーリングを行った．z-score とは，平均を 0，分散を 1 に揃えるスケーリングであり，次式で表される．

$$Z = \frac{X - \mu}{\sigma} \tag{2}$$

ここで，X を各データ値，μ を平均，σ^2 を分散とした．

本研究で用いた訓練データとテストデータの要素数を**表 1** に示す．

表 1 データセットの要素数

	“Active”	“Inactive”	合計
訓練データ	2,477	199	2,676
テストデータ	11	589	600

3.2 評価基準

本研究では精度の評価に ROC 曲線下面積（Area under ROC curve: AUROC）を用いた．また，3 つの予測モデルの比較のために Precision，Recall，F-measure についても確認した．

予測モデルがデータセットに対して予測をしたとき，True Positive となった個数を #TP，True Negative となった個数を #TN，False Positive となった個数を #FP，False Negative となった個数を #FN とする．

- AUROC
  ROC 曲線とは，予測モデルの出力の閾値を変化させながら縦軸に True Positive Rate ($= \frac{\#\mathrm{TP}}{\#\mathrm{TP}+\#\mathrm{FN}}$)，横軸に False Positive Rate ($= \frac{\#\mathrm{FP}}{\#\mathrm{FP}+\#\mathrm{TN}}$) をとった曲線である．ROC 曲線の良さは，ROC 曲線の下側の面積（AUROC）によって評価できる．理想的な予測モデルでは AUROC は 1.0 となり，予測をランダムに行うモデルでは AUROC は 0.5 となる．
- Precision
  Precision とは正例と予測されたものの中で正しく予測できた割合であり，以下の式で表される．
  $$\mathrm{Precision} = \frac{\#\mathrm{TP}}{\#\mathrm{TP} + \#\mathrm{FP}} \tag{3}$$
- Recall
  Recall とは正例のものの中で正しく予測できた割合であり，以下の式で表される．
  $$\mathrm{Recall} = \frac{\#\mathrm{TP}}{\#\mathrm{TP} + \#\mathrm{FN}} \tag{4}$$
- F-measure
  一般に Precision と Recall はトレードオフの関係にある．これら 2 つの指標をまとめた指標として F-measure がある．F-measure は Precision と Recall の調和平均で求められる．
  $$\text{F-measure} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{5}$$

3.3 ハイパーパラメータ探索

本研究では訓練データに対する 5-fold Cross Validation を用いて AUROC が最大となるハイパーパラメータを探索した．3 つの予測モデルそれぞれについてハイパーパラメータの探索方法を以下に示す．

3.3.1 SVM

ガウスカーネルを用いた SVM のハイパーパラメータは，誤分類の許容を調整するコストパラメータ C と決定境界の複雑度を決めるガウスカーネルのパラメータ γ の 2 つがある．これらの 2 つのパラメータを次の手順で決定した．

(1) まず，$C = \{2^{-5}, 2^{-3}, \cdots, 2^{13}, 2^{15}\}$ の 11 通りと $\gamma = \{2^{-15}, 2^{-13}, \cdots, 2^{1}, 2^{3}\}$ の 10 通り，合わせて 110 通りの組み合わせの中から AUROC が最大となるパラメータの組を求める．ここで得られたパラメータの組み合わせを $\{C_0, \gamma_0\}$ とする．

(2) 続いて，$C = \{C_0 \cdot 2^{-2}, C_0 \cdot 2^{-1.75}, \cdots, C_0 \cdot 2^{1.75}, C_0 \cdot 2^{2}\}$ の 17 通りと $\gamma = \{\gamma_0 \cdot 2^{-2}, \gamma_0 \cdot 2^{-1.75}, \cdots, \gamma_0 \cdot 2^{1.75}, \gamma_0 \cdot 2^{2}\}$ の 17 通り，合わせて 289 通りの組み合わせの中から評価指標が最大となるパラメータの組を求める．ここで得られたパラメータの組を最終的な SVM のハイパーパラメータとする．

3.3.2 Deep Learning

Deep Learning による予測モデルには決定すべきハイパーパラメータが複数存在する．また，予測モデルの性能に大きく関わるハイパーパラメータと，そうでないハイパーパラメータがある．Bergstra ら（2012）は Deep Learning におけるハイパーパラメータ探索は，グリッドサーチよりもランダムサーチを用いた方が良いハイパーパラメータを得ることができると報告している [17]．そこで，本研究ではランダムサーチを用いて以下のハイパーパラメータの探索を行った．

- バッチサイズ: $[10 \sim 100]$
- エポック数: $[1 \sim 100]$
- 隠れ層ユニット数: $[128 \sim 4000]$
- 隠れ層の数: $[1, 2, 3]$
- SDA における破壊率: $[0 \sim 1]$
- SDA における正則化パラメータ ρ: $[0 \sim 1]$
- SDA における正則化パラメータ β: $[0 \sim 1]$
- ドロップアウトにおけるユニット選別確率: $[0 \sim 1]$
- fine tuning における正則化パラメータ: $[10^{-5} \sim 10^{-2}]$

3.3.3 Deep Learning + SVM

Deep Learning + SVM による予測モデルでは，以下のハイパーパラメータをランダムサーチによって決定した．

- バッチサイズ: $[10 \sim 100]$
- エポック数: $[1 \sim 100]$
- 隠れ層ユニット数: $[128 \sim 4000]$
- 隠れ層の数: $[1, 2, 3]$
- SDA における破壊率: $[0 \sim 1]$
- SDA における正則化パラメータ ρ: $[0 \sim 1]$
- SDA における正則化パラメータ β: $[0 \sim 1]$
- ドロップアウトにおけるユニット選別確率: $[0 \sim 1]$
- SVM におけるコストパラメータ C: $[2^{-5} \sim 2^{15}]$
- SVM におけるガウスカーネルのパラメータ γ: $[2^{-15} \sim 2^{3}]$

3.4 実験結果

SVM，Deep Learning，Deep Learning + SVM の 3 つのモデルにおいて，訓練データに対する 5-fold Cross Validation によって得られた AUROC，およびテストデータに対する予測によって得られた AUROC，Precision，Recall，F-measure を**表 2** に示す．比較のために，ランダム分類によって得られる予測値も併せて示した．また，テストデータに対する予測によって得られた ROC 曲線を**図 2**，**図 3**，**図 4** に示す．

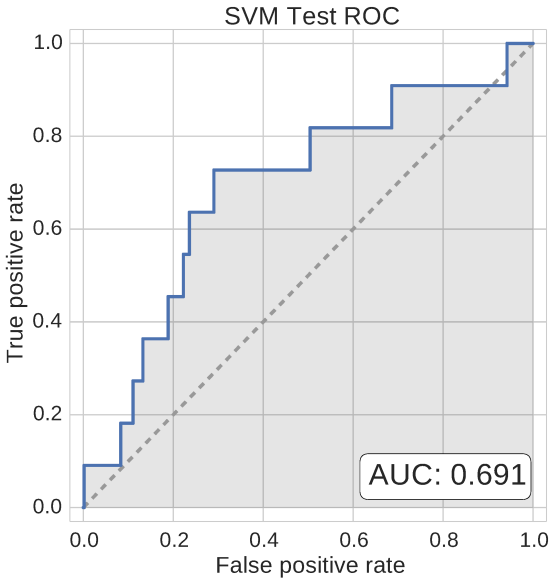

図 2 SVM を用いた予測モデルによって得られたテストデータに対する ROC 曲線

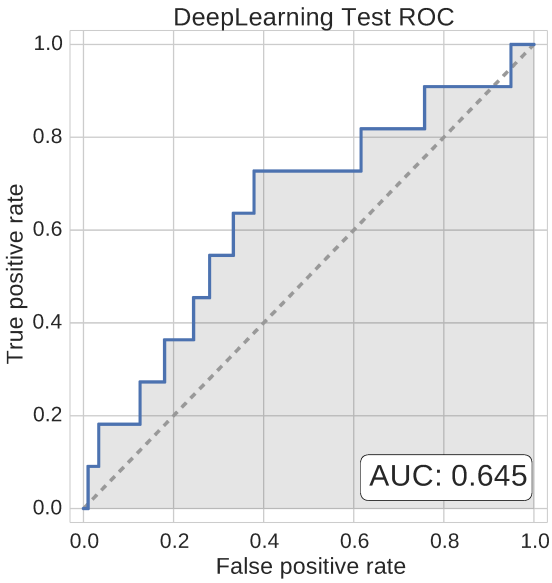

図 3 Deep Learning を用いた予測モデルによって得られたテストデータに対する ROC 曲線

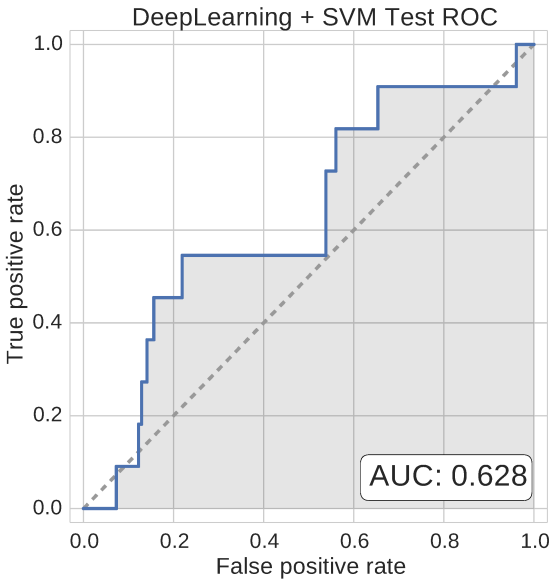

図 4 Deep Learning + SVM を用いた予測モデルによって得られたテストデータに対する ROC 曲線

表 2 訓練データに対する 5-fold Cross Validation，およびテストデータにおける予測の各評価値．表中では Deep Learning を DL と略記した．

	5-fold CV	Test			
	AUROC	AUROC	Precision	Recall	F-measure
SVM	0.987	**0.691**	0.0183	**0.909**	0.0358
DL	0.986	0.645	**0.0199**	**0.909**	**0.0389**
DL+SVM	0.986	0.628	0.0188	**0.909**	0.0368
ランダム分類	0.5	0.5	0.0183	0.5	0.0354

4. 考察

表 2 を見ると，訓練データに対する 5-fold Cross Validation での AUROC は，3 つの予測モデルの間で大差の無い値となっている．テストデータに対する AUROC による評価では SVM による予測モデルが最良であることがわかる．一方で，Deep Learning による予測モデルは AUROC では SVM に劣るものの，Precision と F-measure の値に優れている．Deep Learning + SVM による予測モデルによって得られた AUROC は 3 つのモデルの中で最も低く，Precision と F-measure も Deep Learning による予測モデルに劣っている．

テストデータに対する AUROC の評価で Deep Learning を用いて得られる 2 つの予測モデルが SVM に劣っているのは，訓練データに過学習しているためだと考えられる．kaggle による化合物活性予測コンテスト [4] と本研究の大きな違いは訓練データの要素数である．kaggle による化合物活性予測コンテストの訓練データの要素数はおよそ 15 万に対し，本研究の訓練データの要素数は 2676 である．訓練データの要素数が少ないため，Deep Learning を用いて得られる予測モデルは過学習しているのだと考えられる．

また，3 つの予測モデルによって得られた Recall は全て非常に高い値となったが，Precision はランダム分類によって得られる値と大差の無い値となった．つまり，テストデータの化合物の多くが “Active” と判定されたことを意味している．これは，訓練データの中に含まれている化合物の “Active” と “Inactive” の割合が，“Active” 側に大きく偏っているため，得られる予測モデルの決定境界も “Active” 側に偏ったからだと考えられる．さらに，テストデータの化合物は IPAB コンテストの参加者が阻害活性を持つと予測したものであるため，テストデータの化合物は “Active” と判定されやすいのではないかと考えられる．

テストデータには 11 個の “Active” ラベルが付与された化合物が存在するが，これらに対する 3 つの予測モデルによる予測は全て同じ結果となった．具体的には，id 番号: Z1095352660 を “Inactive” と誤って予測し，他の 10 個の化合物を “Active” と正しく予測した．これらの化合物の構造式を**図 5** に示す．予測結果が True Positive となった化合物分子量の最大値は 462.58，最小値は 252.30，平均値は 385.74 であったのに対し，予測結果が False Negative となった化合物（id 番号: Z1095352660）の分子量は 151.60 であった．この化合物は他の “Active” ラベルが付与された化合物とは異なった特徴を持つため，予測が難しかったのではないかと考えられる．

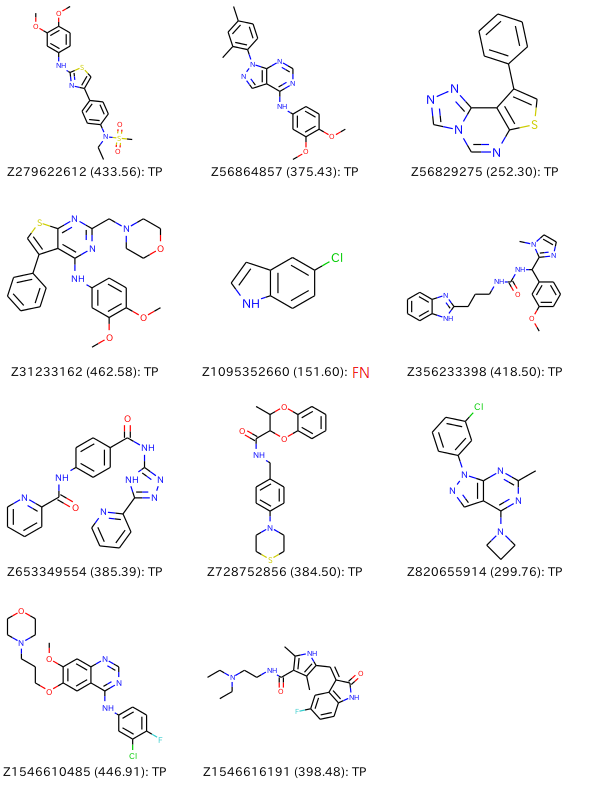

図 5 テストデータの “Active” ラベルが付与された化合物に対する予測結果．各構造式の下部に id 番号，分子量，および予測結果を示している．図中では True Positive を TP と，False Negative を FN と略記した．

5. まとめ

5.1 結論

本研究では，SVM と Deep Learning をそれぞれ用いて標的阻害化合物の予測モデルを構築した．また，Deep Learning の出力層に SVM を用いた予測モデルを構築した．そして，これら 3 つの予測モデルをヒト c-Yes キナーゼ阻害化合物の予測に適用し，精度の評価を行った．

Deep Learning を用いた予測モデルに対して，事前学習や重みの制約などの過学習を避けるテクニックを用いた．また，Deep Learning を用いた予測モデルは決定すべきハイパーパラメータが複数あるため，グリッドサーチではなくランダムサーチを用いて決定した．

テストデータに対する予測によって得られた AUROC が最大となったのは SVM による予測モデルであった．また，3 つの予測モデルに共通して Recall が高い値となった．

テストデータの “Active” ラベルが付与された化合物について，3 つの予測モデルによる予測は全て同じ結果となった．

5.2 今後の課題

5.2.1 Deep Learning に入力する特徴量について

画像認識や音声認識で高い精度を出している Deep Learning には，画像や音声の生データを入力特徴量とすることが多い．本研究では，3 つの予測モデルそれぞれに RDKit で計算される 196 種類の化合物特徴量を入力した．これらの化合物特徴量には分子量や分配係数など複雑な計算の結果として得られる特徴量が含まれている．化合物を部分構造の組み合わせに対応するビット列で表す fingerprint のように，単純な特徴量を Deep Learning の入力として与えることでより高い精度が得られる可能性がある．

5.2.2 ラベルが不均衡な訓練データについて

本研究で用いた訓練データは “Active” ラベルが付与された化合物が 2,477 個，“Inactive” ラベルが付与された化合物が 199 個という不均衡なデータセットであった．Over Sampling を行い “Inactive” ラベルが付与されるサンプルを増やしたり，活性報告のないデータを訓練データに加えて半教師付き学習を行うことで，訓練データの不均衡さを解決できる可能性がある．

謝辞 本研究の一部は日本学術振興会 科研費 基盤研究 (A) (24240044) の支援を受けて行われた．

参考文献

[1] A. Lavecchia, Machine-learning approaches in drug discovery: methods and applications., *Drug Discovery Today*, 20(3): 318–331, 2015.

[2] A. Krizhervsky, I. Sutskever, G. Hinton, ImageNet Classification with Deep Convolutional Neural Networks., In *Proceedings of Neural Information and Processing Systems (NIPS 2012)*, 2012.

[3] A. Graves, A. Mohamed, G. Hinton, Speech Recognition with Deep Recurrent Neural Networks., In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, 6645–6649, 2013.

[4] kaggle; Merck Molecular Activity Challenge, https://www.kaggle.com/c/MerckActivity

[5] A. Hirsch, G.Medigeshi, H. Meyers, *et al.*, The Src family kinase c-Yes is required for maturation of West Nile virus particles., *Journal of Virology*, 79(18): 11943–11951, 2005.

[6] 厚生労働省: ウエストナイル熱・脳炎 Q & A, http://www.mhlw.go.jp/bunya/kenkou/kekkaku-kansenshou08/02.html

[7] C. Ramakrishnan, A. M. Thangakani, D. Velmurugan, *et al.*, Identification of Novel c-Yes Kinase Inhibitors., *Lecture Notes in Computer Science*, 8590: 494–500, 2014.

[8] J. Witek, S. Smusz, K. Rataj, *et al.*, An application of machine learning methods to structural interaction fingerprints-a case study of kinase inhibitors., *Bioorganic & Medicinal Chemistry Letters*, 24(2): 580–585, 2013.

[9] M. Zheng, Z. Liu, X. Yan, *et al.*, LBVS: an online platform for ligand-based virtual screening using publicly accessible databases., *Molecular Diversity*, 18(4): 829–840, 2014.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, Scikit-learn: Machine learning in Python., *Journal of Machine Learning Research*, 12: 2825–2830, 2011.

[11] C. Hsu, C. Chang, C. Lin, A Practical Guide to Support Vector Classification., Tech. rep., 2003.

[12] I. Goodfellow, D. Warde-Farley, P. Lambin, *et al.*, Pylearn2: a machine learning research library., arXiv preprint arXiv:1308.4214, 2013.

[13] D. Erhan, Y. Bengio, A. Courville, *et al.*, Why Does Unsupervised Pre-training Help Deep Learning?, *Journal of Machine Learning Research*, 11: 625–660, 2010.

[14] P. Vincent, H. Larochelle, Y. Bengio, *et al.*, Extracting and Composing Robust Features with Denoising Autoencoders, In *Proceedings of the 25th International Conference on Machine learning (ICML 2008)*, 1096–1103, 2008.

[15] P. Vincent, H. Larochelle, I. Lajoie, *et al.*, Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion., *Journal of Machine Learning Research*, 11: 3371–3408, 2010.

[16] N. Srivastava, G. Hinton, A. Krizhevsky, *et al.*, Dropout : A Simple Way to Prevent Neural Networks from Overfitting., *Journal of Machine Learning Research*, 15, 1929–1958, 2014.

[17] J. Bergstra, Y. Bengio, Random Search for Hyper-Parameter Optimization., *Journal of Machine Learning Research*, 13, 281–305, 2012.

[18] J. Duchi, E. Hazan, Y. Singer, Adaptive Subgradient Methods for Online Learning and Stochastic Optimization., *Journal of Machine Learning Research*, 12, 2121-2159, 2011.

[19] M. Zeiler, ADADELTA: An Adaptive Learning Rate Method., arXiv preprint arXiv:1212.5701, 2012.

[20] Kinase SARfari, European Bioinformatics Institute, https://www.ebi.ac.uk/chembl/sarfari/kinasesarfari

[21] IPAB コンテスト :「コンピュータで薬のタネを創る」, 特定非営利活動法人　並列生物情報処理イニシアティブ, http://www.ipab.org/eventschedule/contest

[22] RDKit, http://www.rdkit.org/