

聖マリアンナ医科大学  
リウマチ・膠原病・アレルギー内科

抄読会

2023.02.27 担当:元永






August 2022

**RMD  
Open**

Rheumatic &  
Musculoskeletal  
Diseases

ORIGINAL RESEARCH

## Machine learning predicts response to TNF inhibitors in rheumatoid arthritis: results on the ESPOIR and ABIRISK cohorts

Vincent Bouget <sup>1</sup>, Julien Duquesne,<sup>1</sup> Signe Hassler,<sup>2,3</sup> Paul-Henry Cournède,<sup>4</sup> Bruno Fautrel <sup>5,6</sup>, Francis Guillemin <sup>7</sup>, Marc Pallardy,<sup>8,9</sup> Philippe Broët,<sup>2,3</sup> Xavier Mariette <sup>10</sup>, Samuel Bitoun <sup>10</sup>

➔ ESPOIRおよびABIRISKのコホートの結果から関節リウマチにおけるTNF阻害剤の反応性を機械学習で予測する。

今話題の機械学習。

リウマチ・膠原病領域でも取り上げられ論文にされるようになってきたので、この題材にしてみました。

RMD(Rheumatic & Musculoskeletal Diseases) Openは、EULAR(European Alliance of Associations for Rheumatology: 欧州リウマチ学会)の公式ジャーナル。IF:5.806。

著者はVincent Bougetさん。医師ではなくエンジニア？。免疫学における人工知能の展開を専門とするSCIENTA Lab (<https://www.scientalab.com/about-us>)という組織を立ち上げ、リウマチ学において共同研究を行っているみたいです。

まず、最初に

- ・AI?、機械学習とは?
- ・機械学習の流れ
- ・今回のクラス分けした手法(モデルと言います)について

# AI?、機械学習とは?

AIは、**Artificial (=人工的な) Intelligence (=知能、知能)**の略。

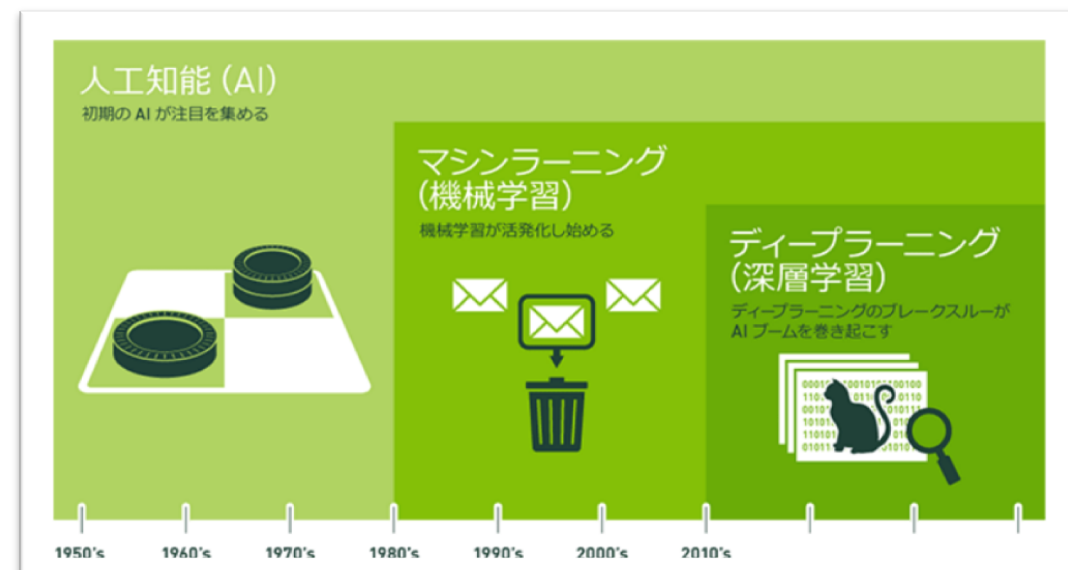
人間のような知能を持つものを人間の手で作る分野として生まれた。

※実際には、コンピュータというデバイスしかない。

インターネットの普及から、サーバーやクラウドなど存在し、

コンピューターも高性能になり扱える情報量が増えた。(もう、人がついていけない)

→ 次は**高度な計算や推論!** (機械学習が生まれた。)

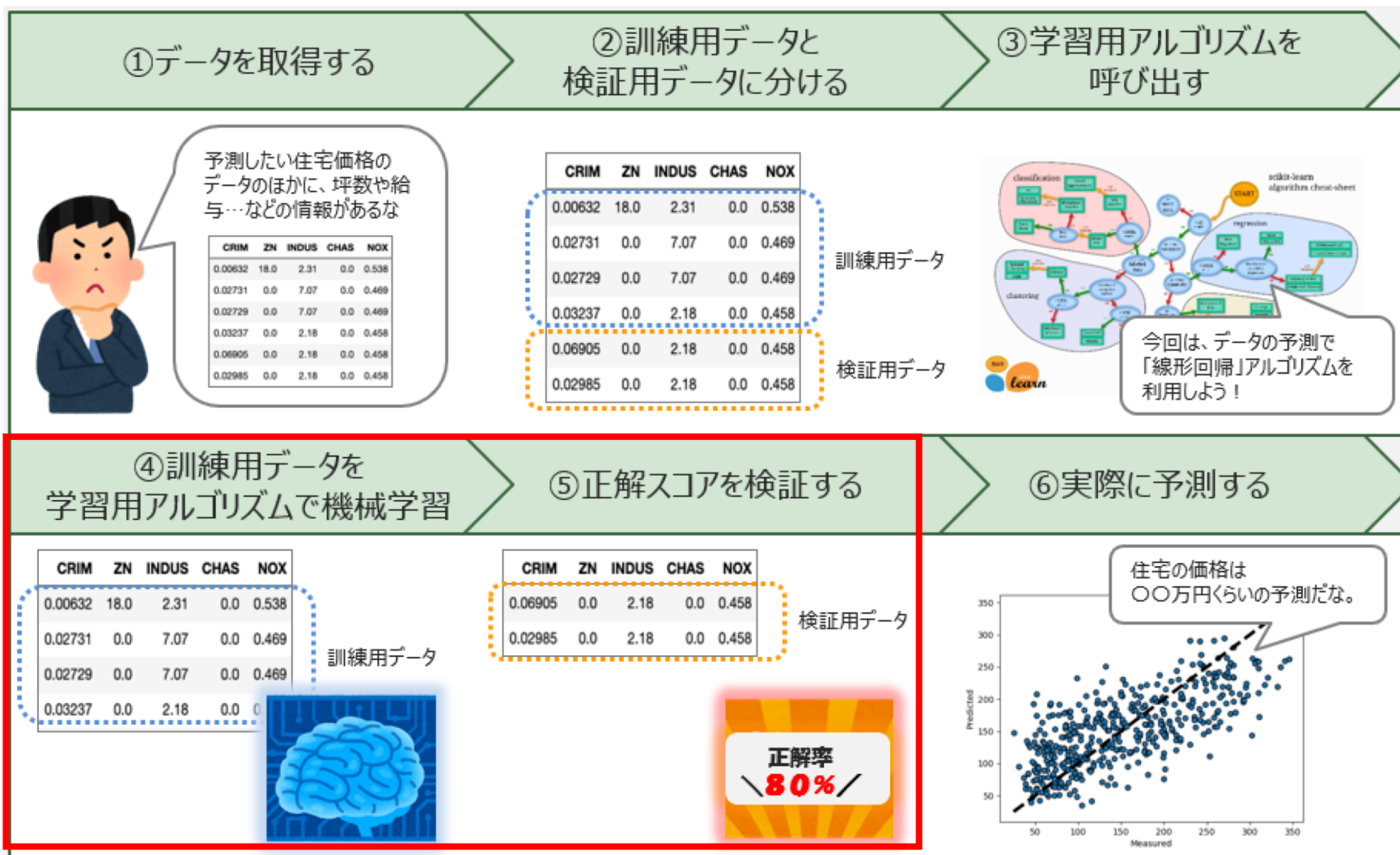


<https://blogs.nvidia.co.jp/2016/08/09/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>

まだまだ、AIに対していろいろな意見があります。

未だ人間が作ったアルゴリズムで推論し、創造性はないという見方もできます。

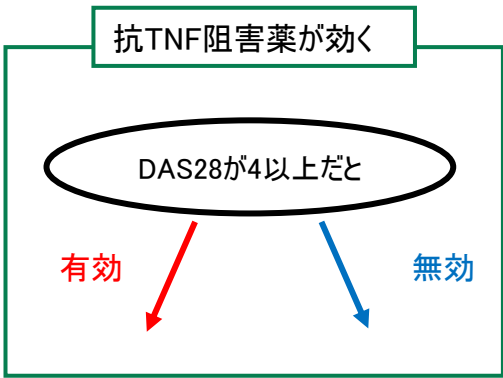
# 機械学習の流れ



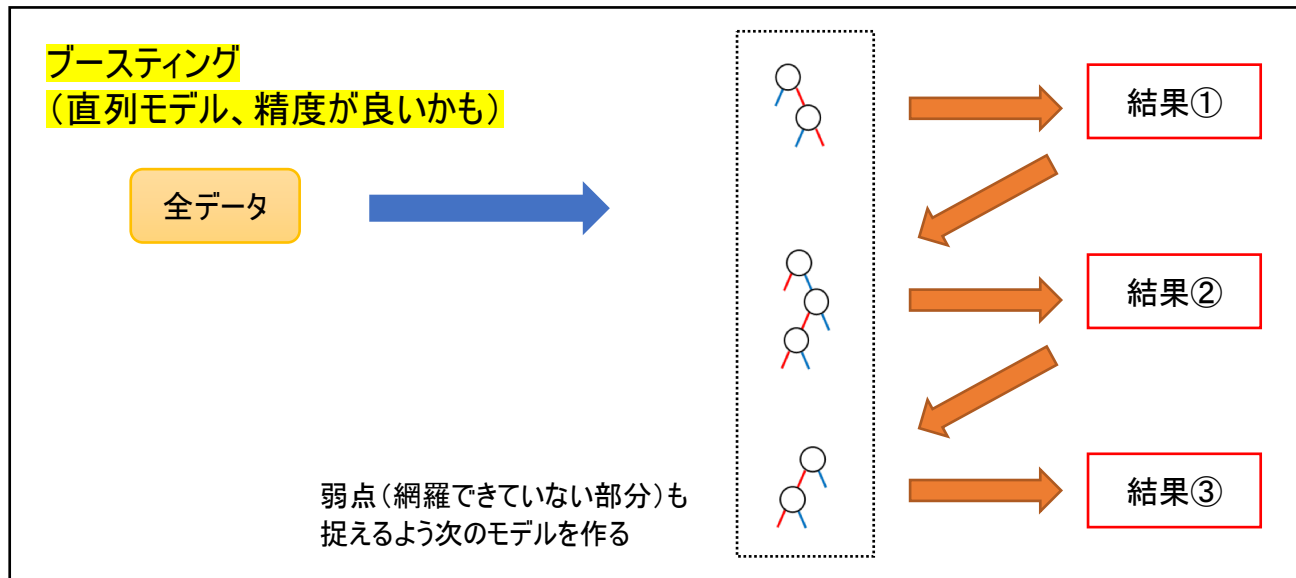
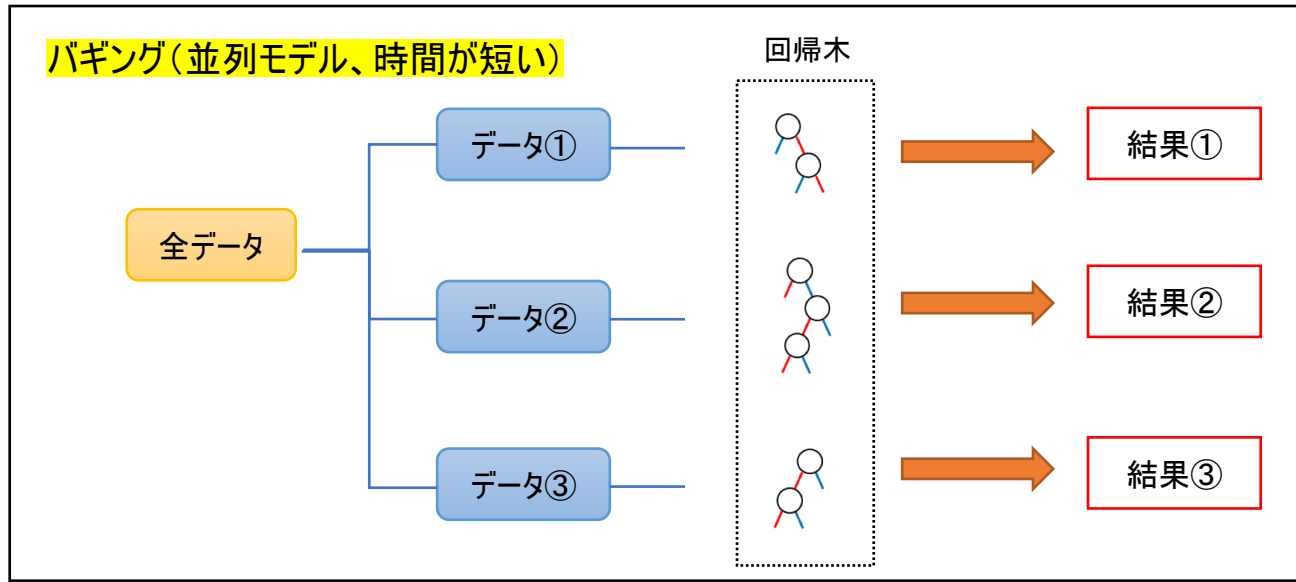
この部分でモデルを作成している。

# 今回のクラス分けした手法(モデルと言います)について

➡一部である決定木モデルのみ



これが決定木。  
さらに、  
有効or無効かという2項なので「分類木」と呼ぶ。  
(数値なら回帰木)



最終的な結果は多数決  
などを行い、導き出す。  
(=アンサンブル学習)

バギング:  
ランダムフォレスト

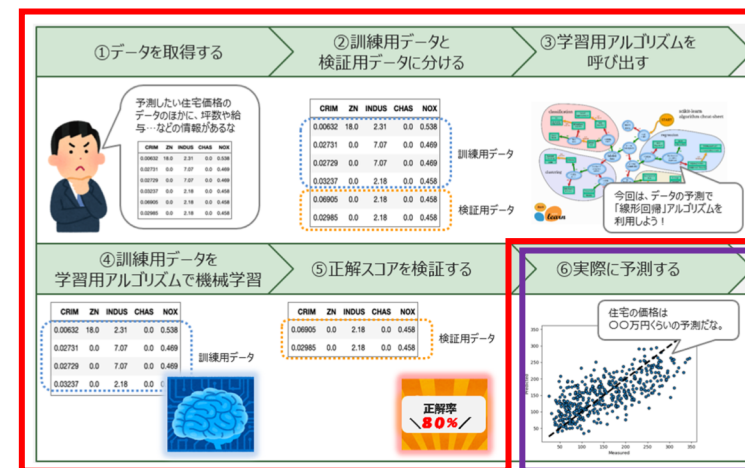
ブースティング: XGBoost,  
CatBoost

## Methods

- ・訓練用データには、2002年に行われたESPOIRコホート試験という関節リウマチ患者のデータを使用した。
- ・治療内容は、①全TNF阻害薬、②エタネルセプト、③抗TNFモノクローナル抗体の群で分類した。
- ・エンドポイントは、治療開始12カ月後のDAS28によるEULARの判定基準を用いた。
- ・検証コホート(実践)は別に用意し、ABIRISKのデータベースを使用した。

※ABIRISKはimi(innovative medicines initiative)が提供しているプロジェクト。  
[ABIRISK | IMI Innovative Medicines Initiative \(europa.eu\)](http://ABIRISK | IMI Innovative Medicines Initiative (europa.eu))

ESPOIR



ABIRISK

- ・モデルの作成には、  
代表的な、線形回帰モデル(ロジスティック回帰)、ランダムフォレスト、XGBoost、CatBoostを使い、  
area under the ROCで比較した。

訓練データ  
これでモデルを作成する。



**Table 1** Characteristics of the training set (ESPOIR) and validation set (ABIRISK) at last visit before treatment initiation

Feature's name	ESPOIR n=161	ABIRISK n=118
Age, years	49 (13)	52 (13)
Female, n (%)	115 (71)	89 (76)
Weight, kg	69 (15)	75 (19)
Height, cm	166 (9)	167 (9)
Body mass index	24.9 (4.6)	25.6 (5.5)
Autoimmunity family history, n (%)	48 (30)	37 (32)
Ever smokers, n (%)	72 (45)	72 (61)
Current smokers, n (%)	28 (17)	33 (28)
Smoking cumulative dose, pack-year	8 (13)	15 (14)
Past pregnancy (among sex=female), n (%)	92 (74)	70 (79)
DAS28	4.6 (1.6)	4.4 (1.1)
CRP, mg/L	17 (27)	12 (16)
Erythrocyte sedimentation rate, mm	26 (23)	23 (19)
Creatinine, μmol/L	74 (17)	66 (14)
AST, U/L	22 (8)	25 (13)
ALT, U/L	22 (13)	27 (20)
White blood, cells/10 <sup>9</sup>	7.9 (2.6)	8.5 (3.2)
Neutrophils, cells/10 <sup>9</sup>	5.4 (2.4)	6.1 (3.2)
Lymphocytes, cells/10 <sup>9</sup>	1.7 (0.68)	2.1 (2.9)
Presence of anti-citrullinated protein antibody, n (%)	113 (70)	81 (70)
Presence of rheumatoid factor IgM	119 (74%)	78 (68%)
TNFi treatment sequences, N	N=208	N=118
Etanercept sequences, N (%)	100 (48)	68 (58)
Monoclonal anti-TNF antibodies sequences, N (%)	108 (52)	50 (42)
Adalimumab sequences, N (%)	80 (74)	39 (78)
Infliximab sequences, N (%)	17 (16)	11 (22)
Certolizumab sequences, N (%)	8 (7)	0 (0)
Golimumab sequences, N (%)	3 (3)	0 (0)
First TNFi line, N (%)	153 (74)	107 (91)
Non-responder imputation, N (%)	10 (4.8)	21 (18)
Responder to sequences, N (%)	122 (59)	72 (61)
Etanercept, N (%)	64 (64)	42 (62)
Monoclonal anti-TNF antibodies, N (%)	58 (54)	30 (60)
Co-treated with corticosteroids, N (%)	94 (45)	51 (43)
Co-treated with MTX, N (%)	152 (73)	64 (54)

Results are presented as follows: mean (SD) for continuous variables and amount (percentage) for binary variables. ALT, aspartate aminotransferase; AST, alanine transaminase; CRP, C reactive protein; DAS28, disease activity score; MTX, methotrexate; TNFi, tumour necrosis factor inhibitors.

## 治療開始する前の最終訪問時のデータ項目

性別、体重、喫煙歴、一般血液検査、DAS28など、どれも普遍的に集まるデータ。

連続変数の平均 (SD) とバイナリ変数の量 (パーセンテージ)。

### エタネルセプト使用症例数(割合)

モノクローナル抗TNF抗体(割合)

### ※注意

ESPOIRに関しては追跡期間中に薬剤変更をすることができた。  
初期にエタネルセプト使用は95名、モノクローナル抗体は96名の合計191だった。合計は161人なので、初期の段階だけでも30人は切り替えられて2剤以上投与されている。

➡同一人物でも、薬剤切り替えタイミングで新規データとしてみるようカウントした。



## Results

**Table 2** Result of the variable selection process for the prediction of the EULAR response

All TNFi	Etanercept	Monoclonal antibodies TNFi
DAS28	DAS28	DAS28
Age	Sex	Sex
Ever smoked	Ever smoked	Ever smoked
Weight	BMI	Weight
Lymphocytes		ESR
Neutrophils		
ALT		

The feature selection was run on the training set (ESPOIR). ALT, alanine transaminase; BMI, body mass index; DAS28, disease activity score; ESR, erythrocyte sedimentation rate; TNFi, tumour necrosis factor inhibitors.

- ・すべてのTNF阻害薬
- ・エタネルセプト
- ・抗TNFモノクローナル抗体

の3つの治療分類において効果予測モデルを作成する上で、より影響する項目(特徴量という)をpick upした。(左の図)。

※残念ながら、その特徴量選定のプロセスについては、

補足資料にしかなく、図はありませんでした。

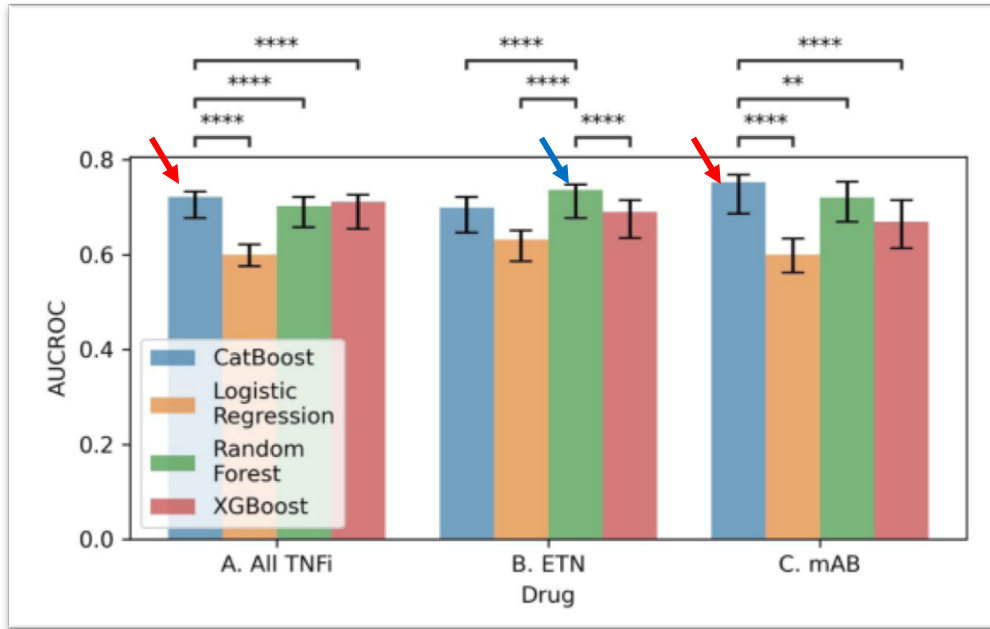
下記を自動化。

- ・トレーニングセットを5パターンに分割してテスト。(5 folds cross-validation)。
- ・すべての項目をランダムフォレストで解析した。
- ・特徴量を逐次削除後に再計算し、影響が小さいものは削除したままに。
- ・モデル機能低下があればそこで終了。

### 【メモ】

ランダムフォレストはバギング。  
並列でモデルを作成し、最後に多数決。  
比較的、多時間で出来る。

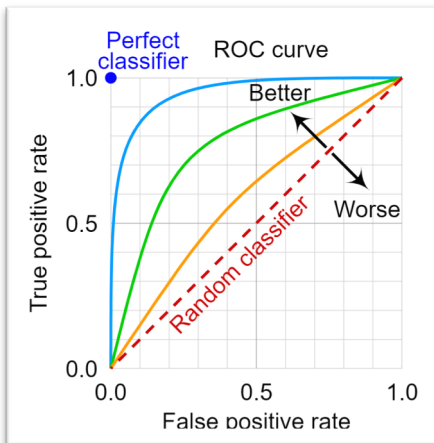
## トレーニングセットでのモデルの評価



- ・すべてのTNF阻害薬
- ・エタネルセプト
- ・抗TNFモノクローナル抗体

3つの群に対して、ロジスティック回帰モデル、ランダムフォレスト、XGBoost、CatBoostという手法のモデルを適応し、AUROCでモデルの評価を行った。

→治療反応を一番予測できたのは**CatBoost**だった。  
エタネルセプトに関しては、**ランダムフォレスト**。



### 【メモ】

- ・AUROC

ROC曲線の下面積。

左の図だと、赤の線はコインを投げて決める方法と同じ精度。

青はパーフェクトな精度を意味する。

面積が大きいと、より青の曲線に近い＝精度が良い とされる。

この部分は欠損値を補完する手段を設定してある。

**Table 3** Performances of the best models for the EULAR response prediction

Drug	Best model	Best missing value imputer	AUROC (train)	AUROC (validation)
Overall TNFi	CatBoost	MICE	0.72 (0.68 to 0.73)	Not evaluated since worse than drug-class-specific models on the training set
Etanercept	Random forest	Median	0.74 (0.68 to 0.75)	0.70 (0.57 to 0.82)
Monoclonal anti-TNF antibodies	CatBoost	MICE	0.74 (0.69 to 0.77)	0.71 (0.55 to 0.86)

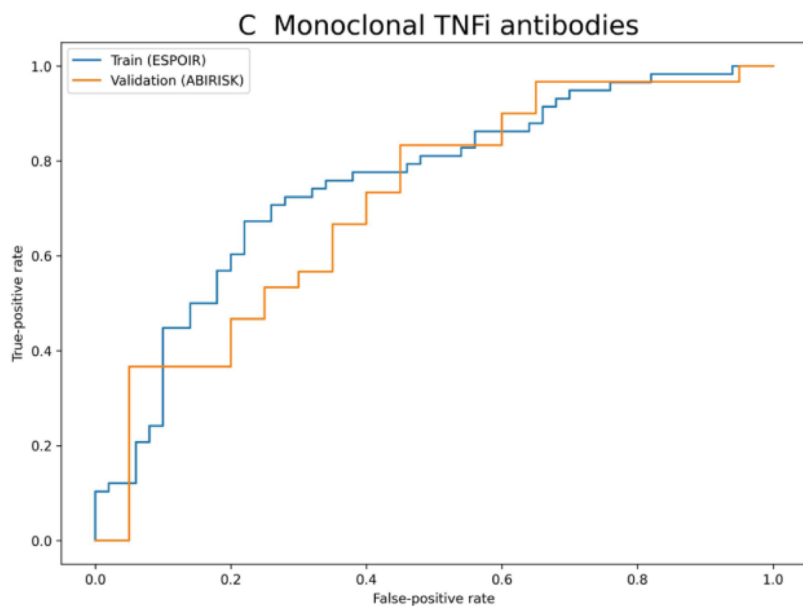
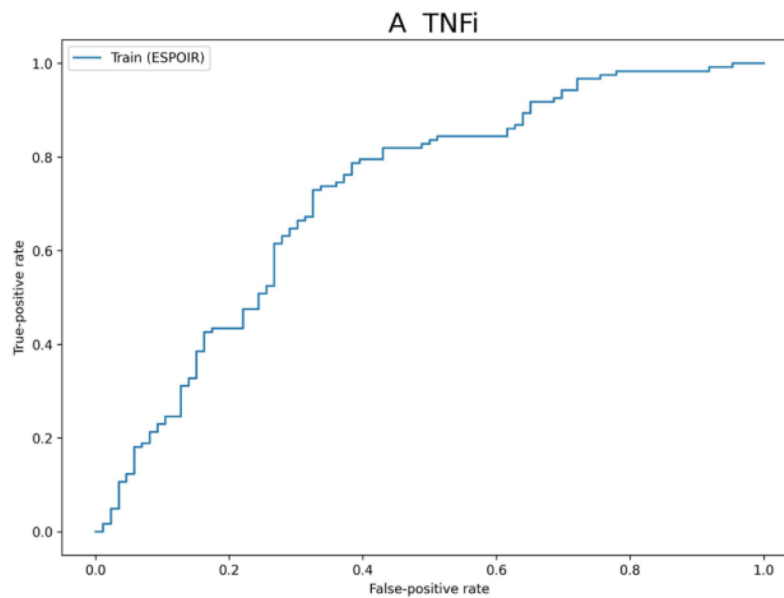
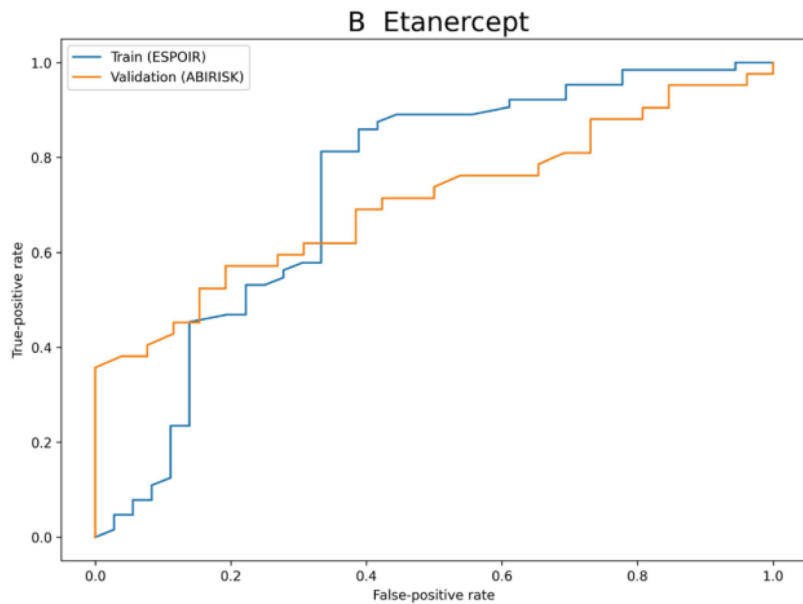
The best model and best missing value imputer were selected on the training set (ESPOIR) using AUROC. The replication of the results was assessed on the validation set (ABIRISK). Numbers in brackets refer to 95% CIs. AUROC, area under the receiver operating characteristic curve; MICE, Multiple Imputation by Chained Equations ; TNFi, tumour necrosis factor inhibitors.

トレーニングセットで作成したモデルを、**検証(validation)セット**で適応してみた。

※トレーニングセットの時点で、全TNF阻害薬で作成したモデルは、エタネルセプトまたは、モノクローナル抗体単独でのモデルに敵わず、検証セットでは評価されなかった。

通常、トレーニングセットで作成したモデルを検証した場合、ほぼ確実に精度は落ちる。

今回は誤差がすくなく、モデルとしても一般化が出来ていると考えられる。



青がトレーニングセットでのROC曲線。

オレンジが検証セットでのROC曲線。

※検証セットで、全TNF阻害のモデルは実施していない。

BとCでROC曲線は比較的一致している。

＝トレーニングセットで作成したモデルは、それ以外の母集団にも使用できる可能性がある。一般化されている。

確実にレスポンドーを探す→陽性的中率:78-92%

確実に不応な人を探す→感度90-95%

**Table 4** Metrics of interest regarding the prediction of the EULAR response

Drug	Strategy 1 (high confidence in response)				Strategy 2 (high confidence in non-response)			
	Sensitivity	Specificity	PPV	NPV	Sensitivity	Specificity	PPV	NPV
Etanercept	60% (44% to 74%)	73% (55% to 89%)	78% (63% to 92%)	53% (36% to 69%)	95% (88% to 100%)	15% (4% to 30%)	64% (52% to 76%)	67% (20% to 100%)
Monoclonal anti-TNF antibodies	37% (20% to 55%)	95% (83% to 100%)	92% (73% to 100%)	50% (35% to 66%)	90% (78% to 100%)	40% (19% to 62%)	69% (54% to 84%)	73% (44% to 100%)

For the two strategies presented in methods, we display the metrics on the validation set (ABIRISK). Numbers in brackets refer to 95% CIs.

NPV, negative predictive value; PPV, positive predictive value; TNF, tumour necrosis factor.

例

Sensitivity	感度	実際に罹患している人の中で、陽性になった人の割合
Specificity	特異度	罹患していない人の中で、陰性になった人の割合
PPV (Positive Predictive Value)	陽性的中率	陽性になった人の中で実際に罹患している人の割合
NPV (Negative Predictive Value)	陰性的中率	陰性になった人の中で罹患していない人の割合

Cut off値を変更することで、より治療効果者の検出高信頼度モデル、もしくは効果不十分者の検出高信頼度モデルを検討した。それぞれマークした部分の確率が上がる。

## Discission

下記、この論文のDiscission一部です。

- ・別コホートで、モデルのAUROCが近似することは、そのモデルが有益であると考ええる。
- ・全TNF阻害でのモデルより、エタネルセプトとモノクローナル抗体をそれぞれモデル作成した方が結果が良かった。それらの薬剤が別の機序で作用し、変数も異なる可能性、異なる重みの可能性がある。
- ・別の研究で、ゲノムデータを使用したモデルがあったが、それはAUROC 0.62であった (Guan Y, et.al, 2019)。
- ・モデルの閾値を調整することで、目的に合わせて実臨床に適応できる。元々の文献では治療奏効率は60～65%程度とされているが、80～90%の信頼度でレスポンドーを選定することができる。(それ以外はエタネルセプトとモノクローナル抗体以外で治療すればよい。)

## 【補足】

論文中では、DAS28を予測するモデル作成や、説明変数の項目（特徴量）がモデルにどれくらい貢献しているかを視覚化する「シャープレイ値」という手法についても述べられていましたが、時間上の問題で割愛致しています。