

PCクラスタワークショップ in 大阪2023「ビッグデータとHPC」

## ヘルスケアデータ解析への説明可能AIの応用

2023/06/21

株式会社日立製作所 インダストリアルデジタルビジネスユニット  
エンタープライズソリューション事業部 医薬システム本部

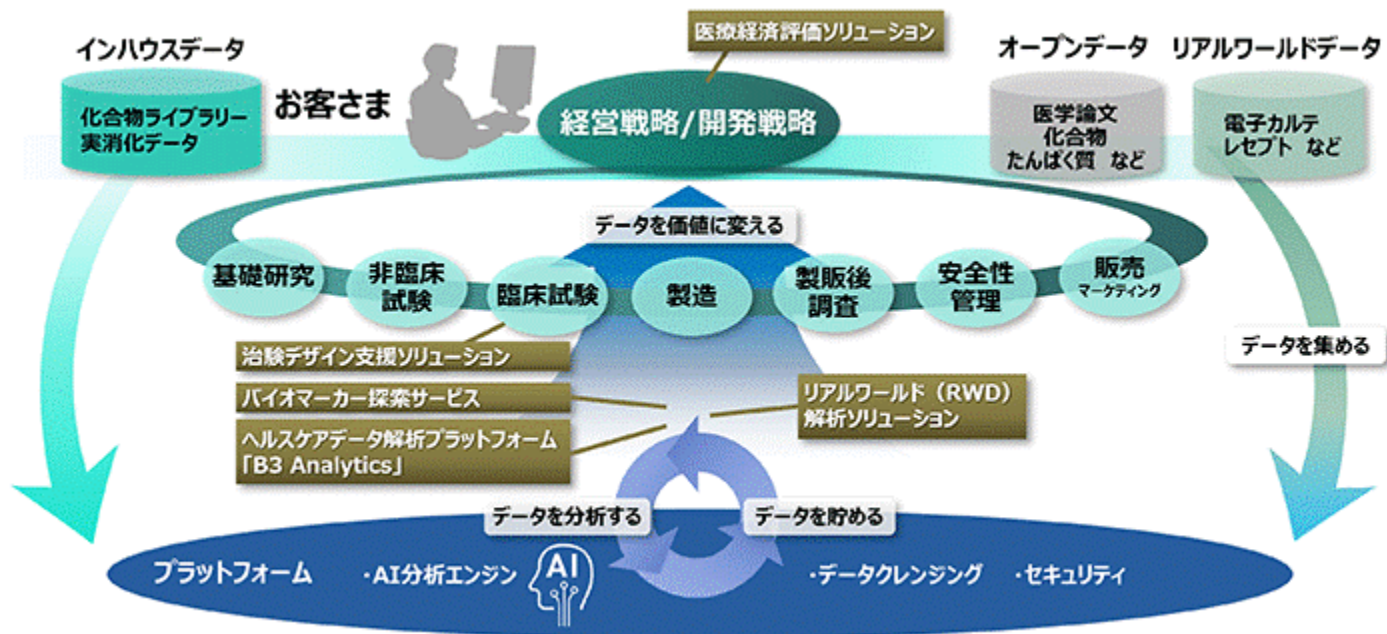
根本 翔太

## Contents

---

1. ヘルスケアデータ解析で求められるAI
2. 日立独自説明可能AIのご紹介
3. ヘルスケアデータへの日立独自説明可能AIの適用事例
4. 今後の展望
5. まとめ

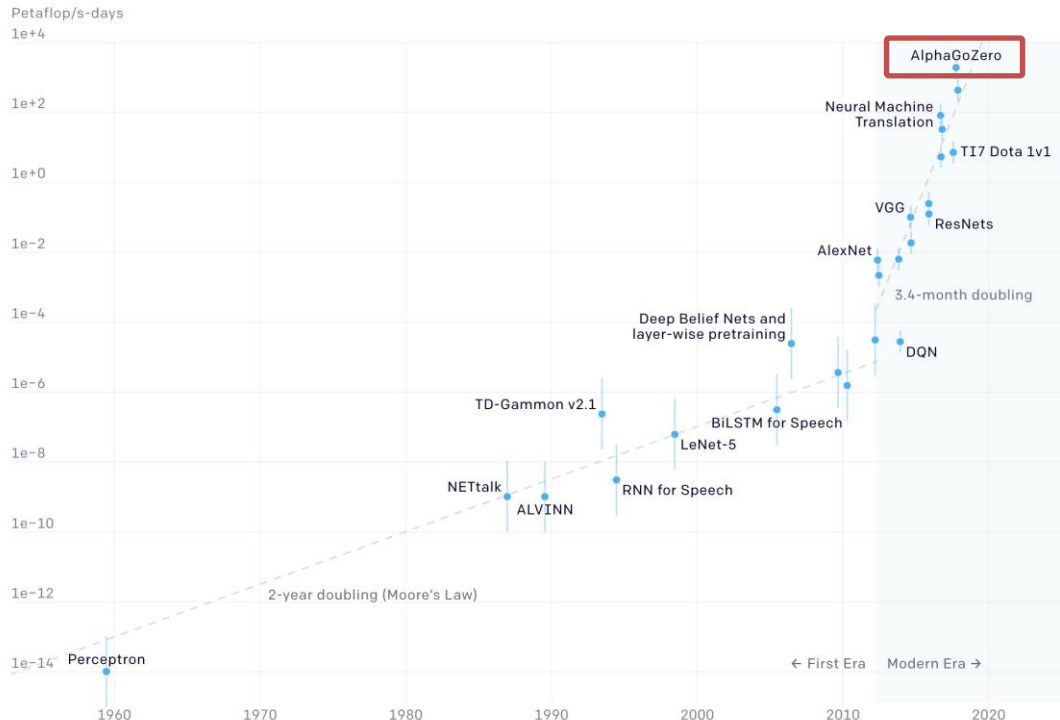
上流の経営戦略/開発戦略から下流の販売/マーケティングまで、創薬バリューチェーンにおける業務プロセスの効率化や精度向上をめざしたサービスを提供します



# 1-1. データ解析のパラダイム

## 演算処理の高速化は、AI技術の進展を支えてきた

Two Distinct Eras of Compute Usage in Training AI Systems



Ref: ARK Investment Management LLC, "AI and Compute." OpenAI, <https://arkinv.st/2ZOH2Rr>.

近年、AIのトレーニングモデルに費やされる  
コンピューターの演算処理能力は急激に増加

➤ AlphaGoZero : 1日当たり1秒間に1.0E+4  
petaflops



AI技術の進展と演算処理能力の  
高速化は密接に関連している

※ AlphaGo Zeroは、DeepMind Technologies, Ltd. の  
日本における登録商標です

## ヘルスケアのあらゆる分野でAIの活用は進んでいる

### 医療・ヘルスケアのあらゆる分野に AI・ICT を応用



## ヘルスケアのあらゆる分野でAIの活用は進んでいる

### 医療・ヘルスケアのあらゆる分野

AIの結果を患者に反映するには  
医師による意思決定が必要

健康維持

病気予防

病気治療

回復・リハビリ

ヘルスケア・未病ケア

健康データ, 生活習慣データ  
ヘルスケア, 各種センサー

AI



オンライン診療  
つながる患者  
妊婦ケア  
...

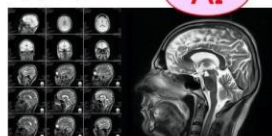
検査

血液, 他  
CT, MRI  
X線, エコー  
ゲノム, ...

診断

診断  
支援

AI



5G 遠隔医療



投薬

薬剤選定  
治療計画

がんゲノム  
医療

新薬の開発

手術

手術支援  
ロボット

論文 + 症例  
臨床試験

新しい医療機器  
の開発



回復ケア・介護

慢性, アドヒアランス  
回復ケア, 自立支援  
オンライン診療  
リハビリ  
介護  
...



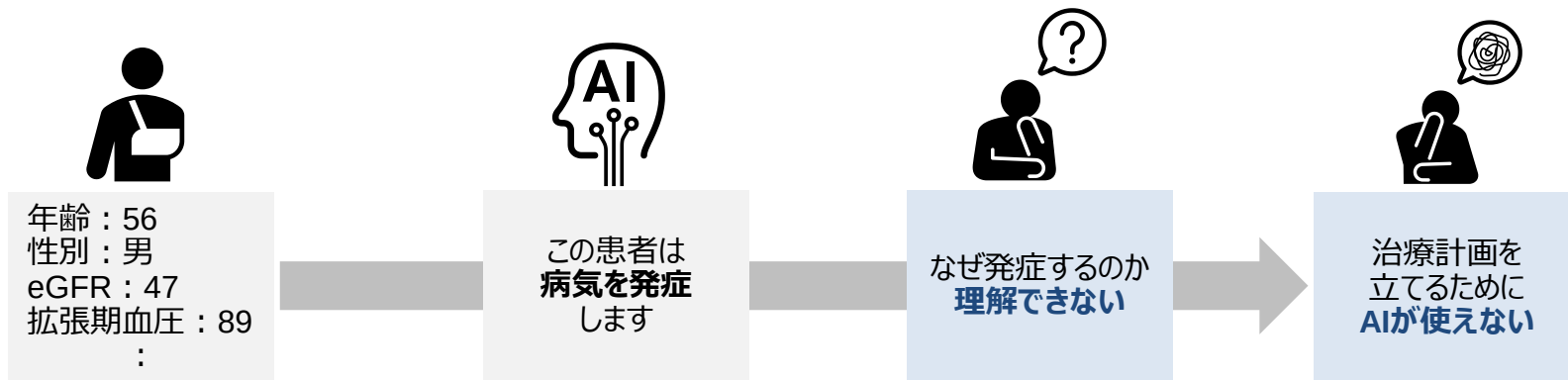
医療ビッグデータ分析

レセプト カルテ 処方箋 ...

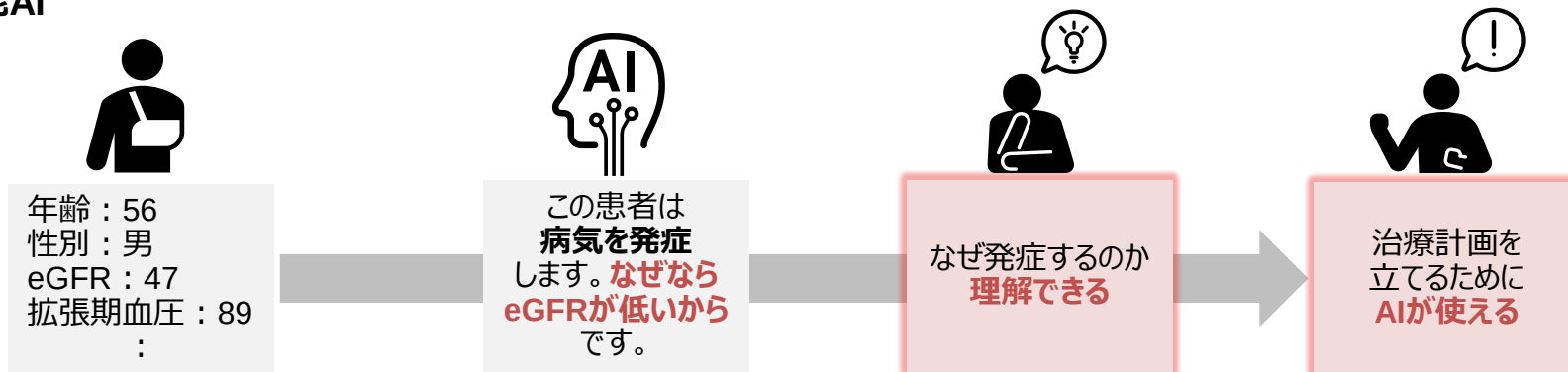
AI

### AIの解析結果をもとに医師が意思決定をする際には“根拠”の説明が必要

#### 従来AI



#### 説明可能AI

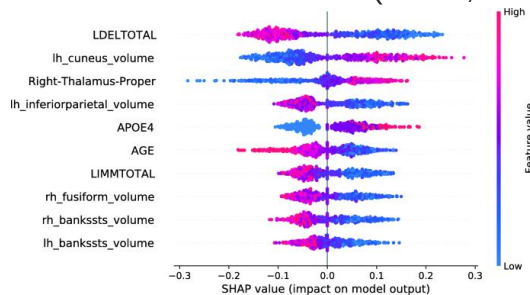


# 1-3. ヘルスケアデータ解析で求められる“説明性”とは

## さまざまな説明方法が開発されているが、どの説明方法がヘルスケア領域と親和性が高い？

### Surrogate Model

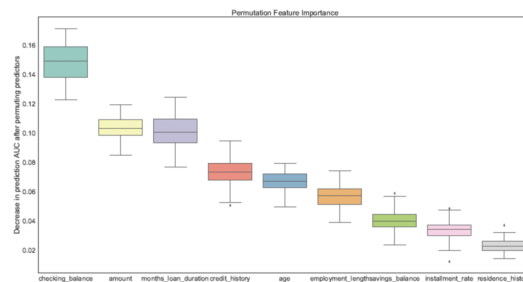
近似モデルを用いて説明を試みる (SHAP, LIME等)



Bloch, L., et al. *Alz Res Therapy* 13, 155 (2021).

### Global Model-Agnostic Methods

特徴量と予測値の関係性を提示



Klosok, Marta & Chlebus, Marcin. 10.13140/RG.2.2.25170.79041. (2020).

### Saliency Map

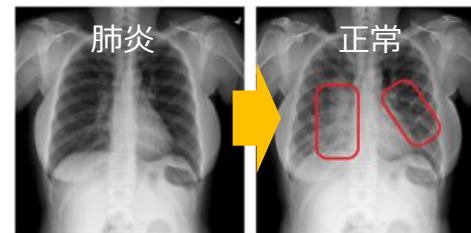
どの領域が予測に影響を与えているのかを提示



Zhang, Qinglong, et al. arXiv preprint arXiv:2103.13859 (2021).

### Counterfactual Explanations

“特徴量の値が今と違ったらどうなるか”を仮想的なサンプルで例示



Mertes, Silvan et al. (2022). *Frontiers in Artificial Intelligence*. 5. 10.3389/frai.2022.825565.

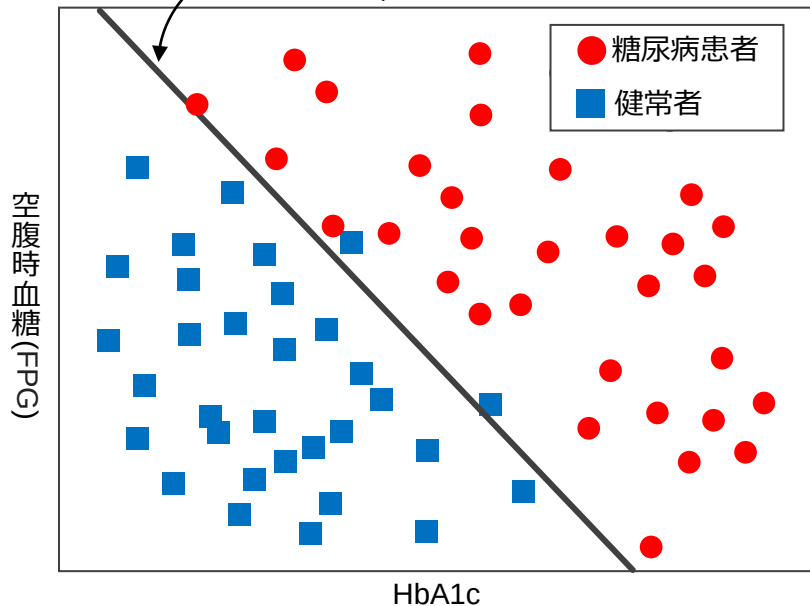
# 1-3. ヘルスケアデータ解析で求められる“説明性”とは

## ヘルスケアデータ解析では“ロジスティック回帰”の説明方法がデファクト

### ロジスティック回帰

- 目的変数（応答変数）が0と1からなる2値のデータについて、説明変数を使った式で表す方法
- 一般化線形モデルの一種であり、空間上を直線で層別（線形識別）
- 説明変数に対する重み（回帰係数）を見ることで、因子の重要度を議論可能

$$y = \sigma(w_1 \times (HbA1c) + w_2 \times (FPG)) \quad (y: \text{予測値}, w: \text{回帰係数})$$



### ロジスティック回帰の出力

| 説明変数  | 回帰係数 | p-value |
|-------|------|---------|
| HbA1c | 3.56 | 0.001   |
| FPG   | 2.96 | 0.007   |

因子の重要度を議論するための統計量

- 回帰係数：回帰係数が大きければ、それに対応する因子が重要
- P-value：回帰係数の有意差検定

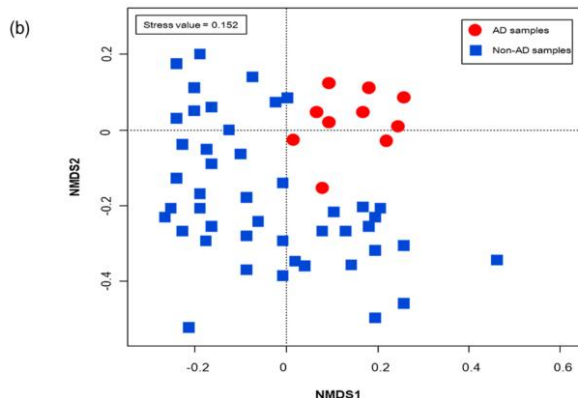
統計量があれば、有効性を定量的に議論できる

# 1-3. ヘルスケアデータ解析で求められる“説明性”とは

現実にはロジスティック回帰では十分に解けない（＝非線形な）問題が多い

## Case 1: 有害事象

右上の領域のみに有害事象例がエンリッチ  
(有害事象の発生には非線形性がある)



Ref: [Classification and Regression Tree Approach for Prediction of Potential Hazards of Urban Airborne Bacteria during Asian Dust Events | Scientific Reports \(nature.com\)](#)

## Case 2: 腎機能の推定

| eGFR 値          | ≥ 90             | 60 ~ 89       | 45 ~ 59               | 30 ~ 44              | 15 ~ 29       | < 15           |
|-----------------|------------------|---------------|-----------------------|----------------------|---------------|----------------|
| 病期              | 第 1 期<br>正常～軽度低下 | 第 2 期<br>軽度低下 | 第 3 期 (前期)<br>軽～中程度低下 | 第 3 期 (後期)<br>中～高度低下 | 第 4 期<br>高度低下 | 第 5 期<br>末期腎不全 |
| 診療ガイドでの治療方針     |                  |               | 専門医へ紹介                | 経過観察                 |               | 透析準備<br>透析     |
| 腎臓病治療の現状 (協会調べ) | 経過観察             |               |                       | 専門医へ紹介               | 透析準備<br>透析    | 透析             |

日本腎臓学会「CKD 診療ガイド 2012」より

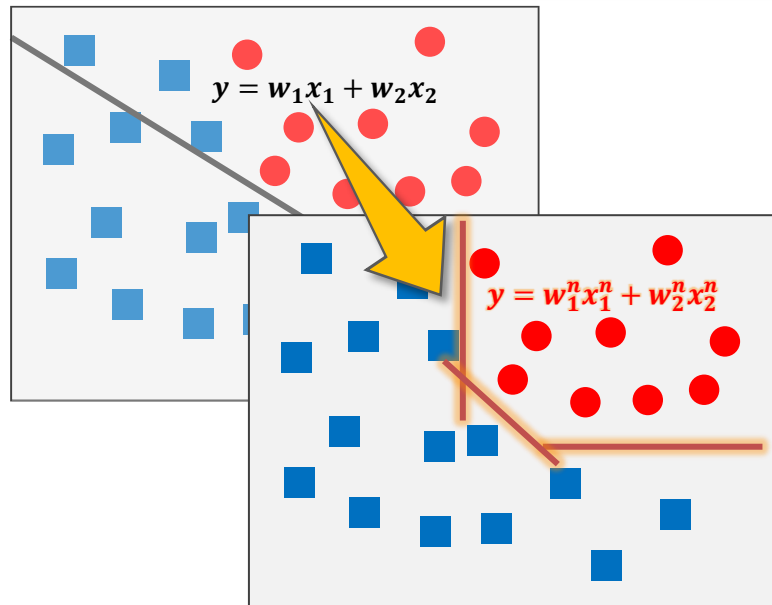
Ref: [自分で知る腎臓健康度チェック | 自分で腎臓病を治す、クレアチンを下げるための支援を行う内臓トレーニング協会 \(naizou.jp\)](#)

$$\begin{aligned} \text{eGFR (mL/min/1.73 m}^2\text{)} \\ &= 175 \times (\text{Scr})^{-1.154} \times (\text{Age})^{-0.203} \\ &\quad \times (0.742 \text{ if female}) \\ &\quad \times (1.212 \text{ if African American}) \end{aligned}$$

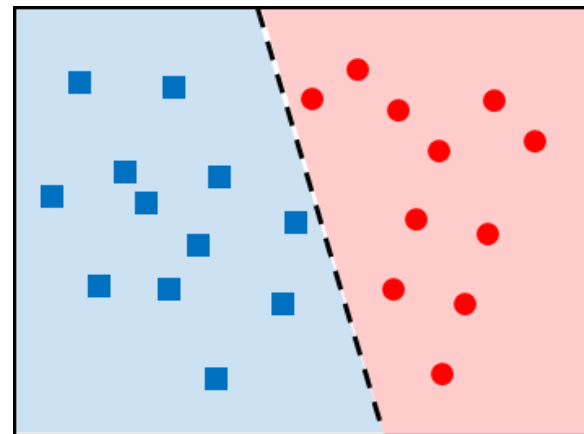
生体メカニズムは複雑なので、 $y = w_1 x_1$  のように線形な関係で説明できる事象は多くない  
→ “ロジスティック回帰”を応用して非線形な関係を説明することはできないか？

ロジスティック回帰よりも高い予測性能をもち、かつヘルスケアデータ解析でデファクトとして使われる“ロジスティック回帰”と同じような解釈が可能なAIを開発

ロジスティック回帰を用いて複数の直線を引く  
(非線形な事象を“直線”で説明する)



ロジスティック回帰で解ける“因子”を生成  
(eGFRのような“新たな因子”を作り出す)

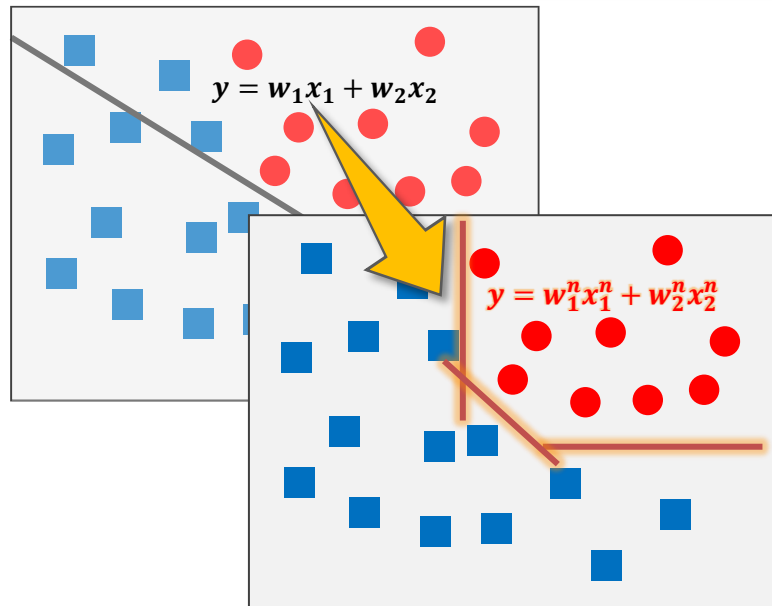


$\frac{\text{height}}{(\text{weight})^2}$

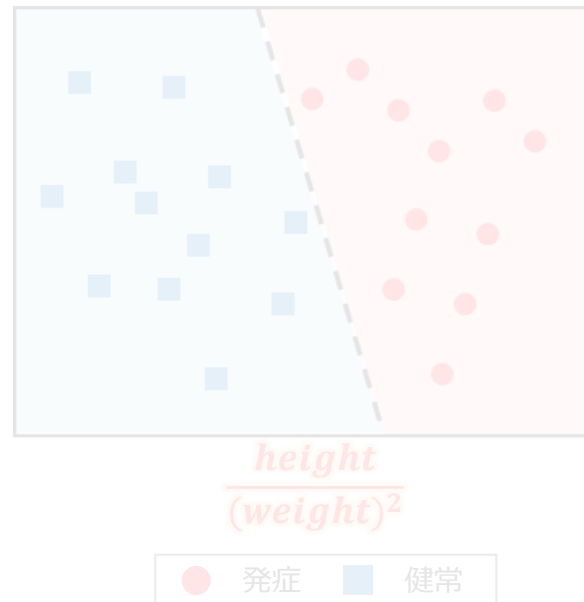
● 発症    ■ 健常

ロジスティック回帰よりも高い予測性能をもち、かつヘルスケアデータ解析でデファクトとして使われる”ロジスティック回帰”と同じような解釈が可能なAIを開発

ロジスティック回帰を用いて複数の直線を引く  
(非線形な事象を”直線”で説明する)

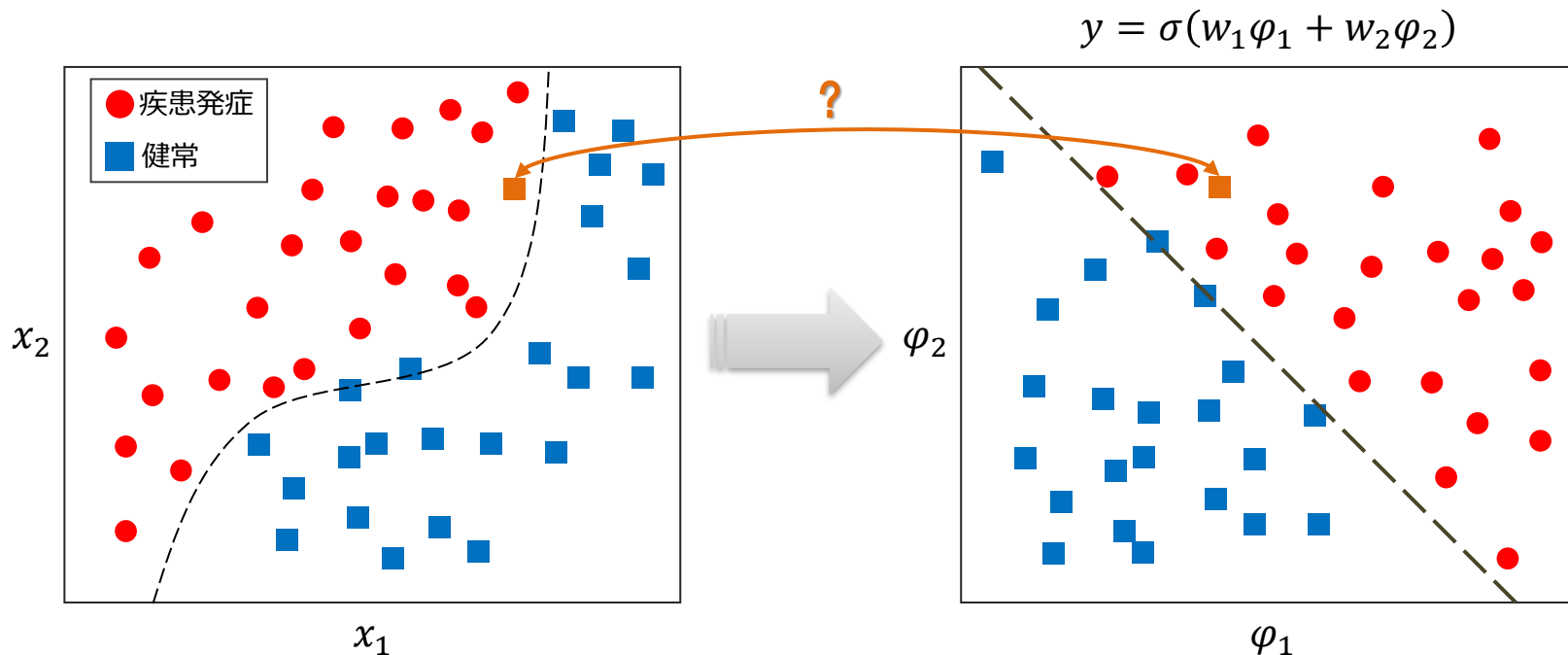


ロジスティック回帰で解ける”因子”を生成  
(eGFRのような”新たな因子”を作り出す)



### 一般的なdeep learning

- 線形分離可能な因子を作り出して（因子を再配置して）、予測モデルを生成
- 新たに生成された因子（ $\varphi_1, \varphi_2$ ）と入力因子（ $x_1, x_2$ ）の関係は不明なため、入力因子に対してロジスティック回帰のような解釈を加えることは不可能

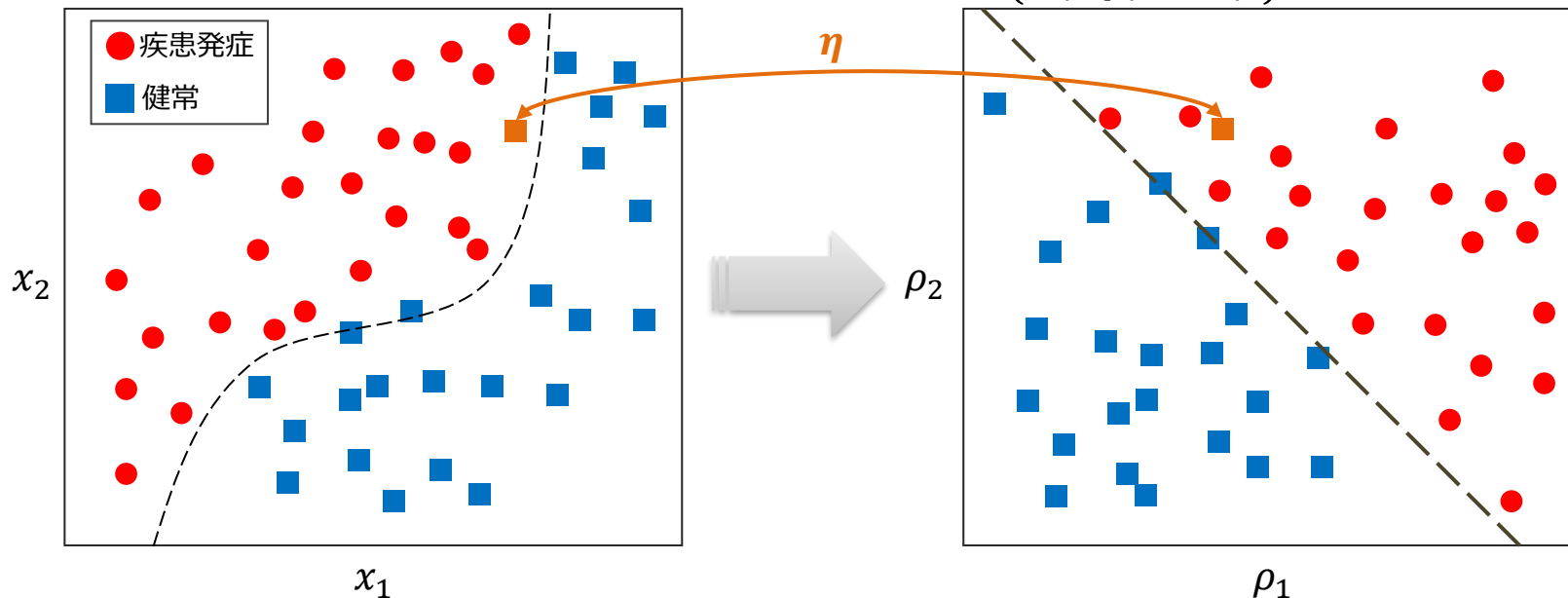


## 2-1. Point-wise linear model : B3

### 日立独自の説明可能AI : B3 (Point-wise linear model)

- transfer vector (reallocation vector)  $\eta$  を保持しており、元の空間に還元することが可能
  - 新たに生成された因子 ( $\rho_1, \rho_2$ ) と入力因子 ( $x_1, x_2$ ) の関係を理解可能

$$\begin{aligned} y &= \sigma(w_1\rho_1 + w_2\rho_2) = \sigma(w \odot \eta(x_n)) \\ &= \sigma(w_n(\eta_n \odot x_n)) \end{aligned}$$



## 2-1. Point-wise linear model : B3

- B3モデルは線形モデルと同様に“回帰係数”を見ることによって、ロジスティック回帰と同様に出力の根拠を理解することが可能

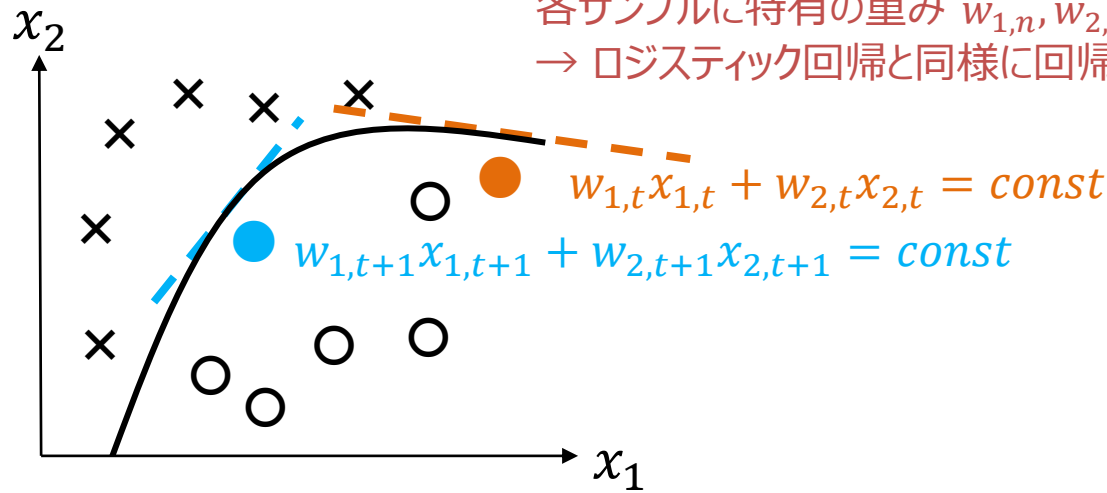
- B3モデルと線形モデルの違い

サンプル  $n$  のもつ因子  $x_{1,n}, x_{2,n}, \dots, x_{D,n}$  に対して、その出力  $y_n$  は

ロジスティック回帰モデル :  $y_n = \sigma(w_1 x_{1,n} + w_2 x_{2,n} + \dots + w_D x_{D,n})$

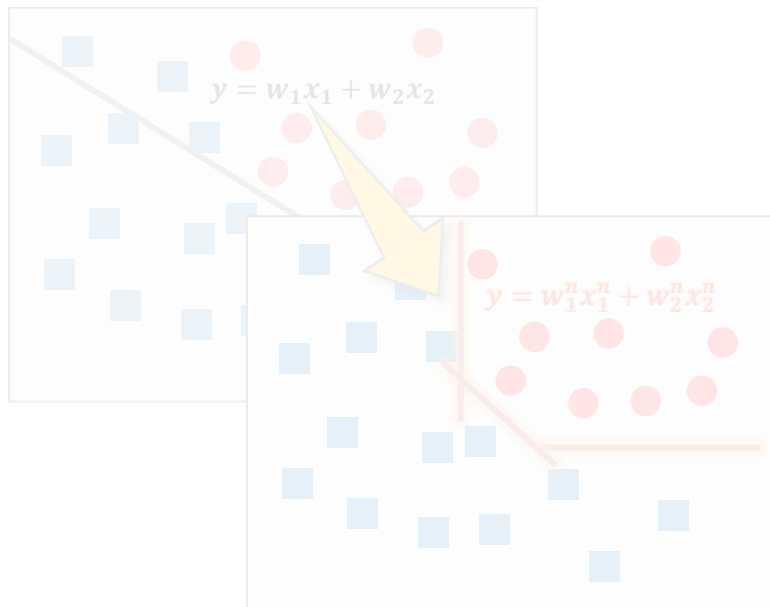
B3モデル :  $y_n = \sigma(\underline{w_{1,n}} x_{1,n} + \underline{w_{2,n}} x_{2,n} + \dots + \underline{w_{D,n}} x_{D,n})$

各サンプルに特有の重み  $w_{1,n}, w_{2,n}, \dots, w_{D,n}$  を計算  
→ ロジスティック回帰と同様に回帰係数で議論できる

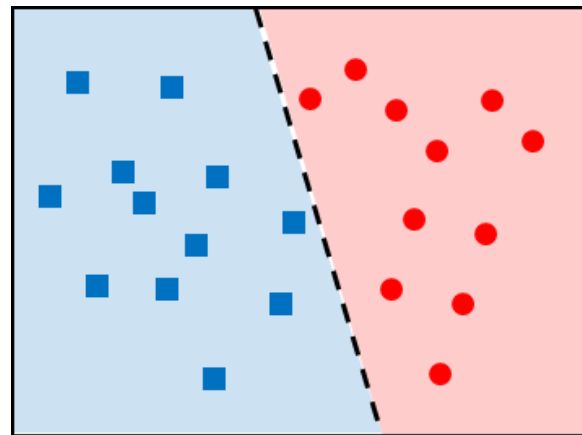


ロジスティック回帰よりも高い予測性能をもち、かつヘルスケアデータ解析でデファクトとして使われる”ロジスティック回帰”と同じような解釈が可能なAIを開発

ロジスティック回帰を用いて複数の直線を引く  
(非線形な事象を”直線”で説明する)



ロジスティック回帰で解ける”因子”を生成  
(eGFRのような”新たな因子”を作り出す)

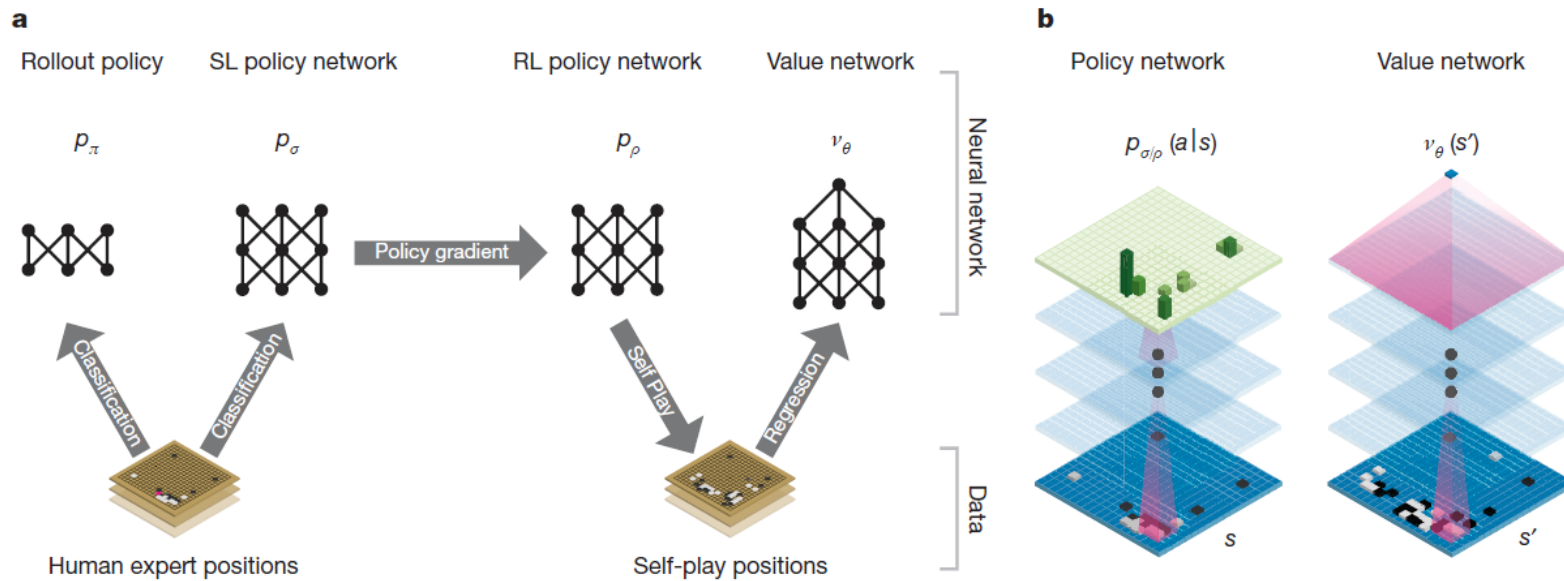


$\frac{\text{height}}{(\text{weight})^2}$

● 発症    ■ 健常

### 強化学習の導入により、効率的に因子を生成できないか？

- AlphaGo : AIが自ら学習し、人間が碁を打つのとまったく同じように振る舞う
  - 同様の仕組みで有用な因子を生成できないか？



Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search." **Nature** 529.7587 (2016): 484.

ADA: 画像(散布図)をみながら、次の一手(座標軸の数式)を考案する

アプローチ 画像を参照しながら数式を解く！

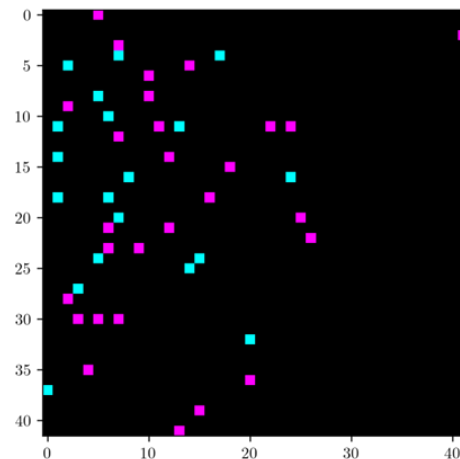
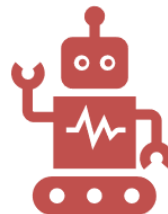
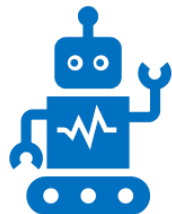
アクション：数式を入力

1. 数式立案

2. 識別器



データと理解可能な演算の  
種類を入力  
Ex. 1軸 2変数まで  
冪乗は2次まで



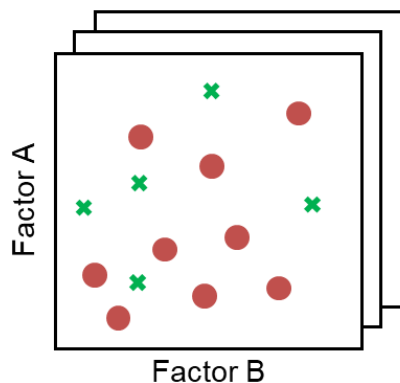
報酬：予測精度

環境：画像を見る

## 2-2. 因子生成AI : ADA

- 強化学習を導入することで、人間と同じようにバイオマーカー候補となる数式を探索する。
- AIは、奏効・非奏効の識別性の高さ(予測精度)を最大化するように学習を行う。

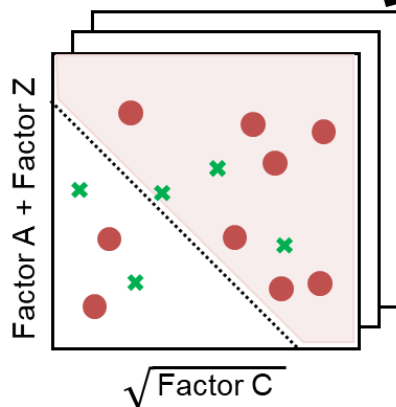
2次元プロットでは  
因子間の関係性が不明瞭



- 探索対象因子、および演算子を予め指定
- アウトプットする予測精度を指定



予測精度を報酬として  
変数への演算を学習

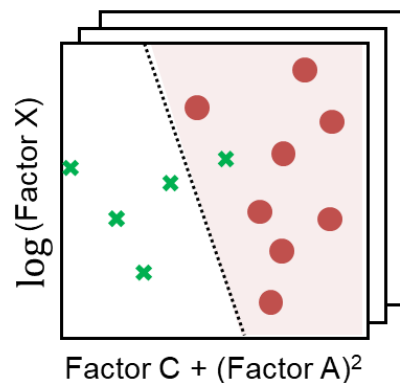


- 指定された因子、演算子から、報酬を最大化する数式を探索



強化  
学習

高い予測精度を  
達成した指標をアウトプット



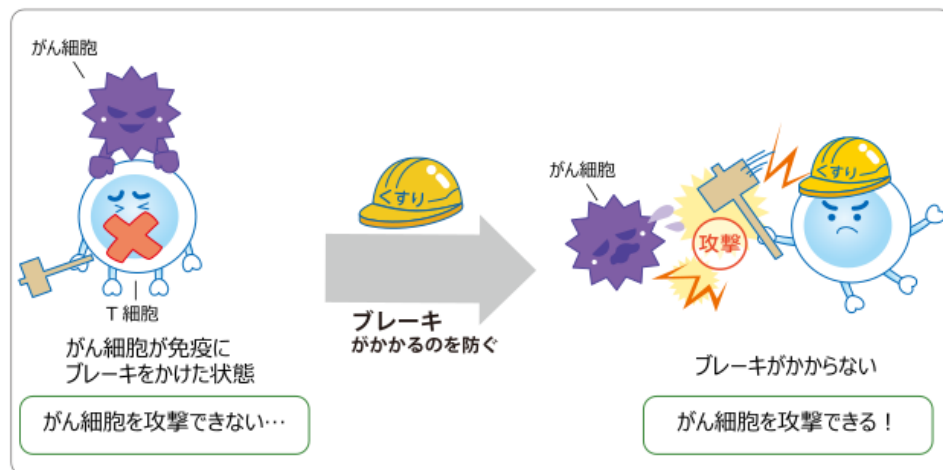
- 指定された予測精度を超えるモデルが生成された場合、画像をアウトプット



## 課題：免疫チェックポイント阻害剤の高性能な治療効果予測

### 免疫チェックポイント阻害剤

- 免疫ががん細胞を攻撃する力を保つ薬
  - 薬剤非投与時：T細胞にブレーキがかかり、がん細胞は排除されなくなる
  - 薬剤投与時：T細胞やがん細胞のアンテナに作用して、免疫にブレーキがかかるのを防ぐ
- 複数の製薬会社が開発に参入



| 薬の種類 ※1   | 薬の名前 ※2          |
|-----------|------------------|
| PD-1阻害薬   | ニボルマブ（オプジーボ）     |
|           | ペムブロリズマブ（キイトルーダ） |
| CTLA-4阻害薬 | イビリムマブ（ヤーボイ）     |
| PD-L1阻害薬  | デュルバルマブ（イミフィンジ）  |
|           | アテゾリズマブ（テセントリク）  |
|           | アベルマブ（バベンチオ）     |

※1 T細胞やがん細胞の表面にあるアンテナを標的にした薬で、標的にするアンテナによって分類します。アンテナの種類にはPD-1、CTLA-4、PD-L1などがあります。

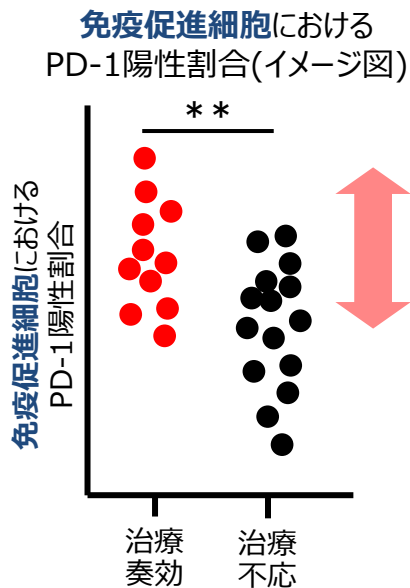
※2 一般名（商品名）を示しています。薬によって、使用できるがんの種類が異なります。

免疫療法 もっと詳しく：[国立がん研究センター がん情報サービス 一般の方へ] ([ganjoho.jp](https://ganjoho.jp))

### 3-1. ヘルスケアデータへの日立独自AI適用事例①

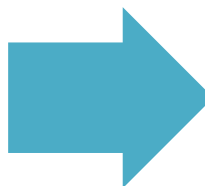
単一因子では高精度に免疫チェックポイント阻害剤の治療効果を予測できないため、  
**因子の組み合わせを考慮したバイオマーカー探索により高精度化を試みた**

#### 単一因子：免疫促進細胞のみ



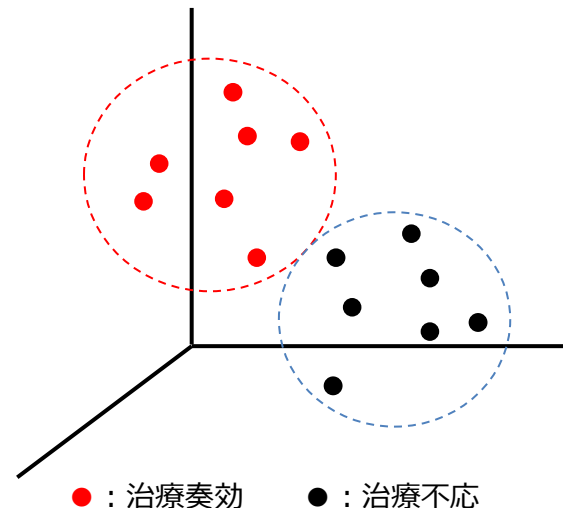
課題：  
統計的有意差は  
あるが、奏効・不応  
の判別が不十分

複数因子



極力、少ない  
因子の組み  
合わせ

#### 複数因子の組み合わせ： 免疫促進細胞、免疫抑制細胞、...

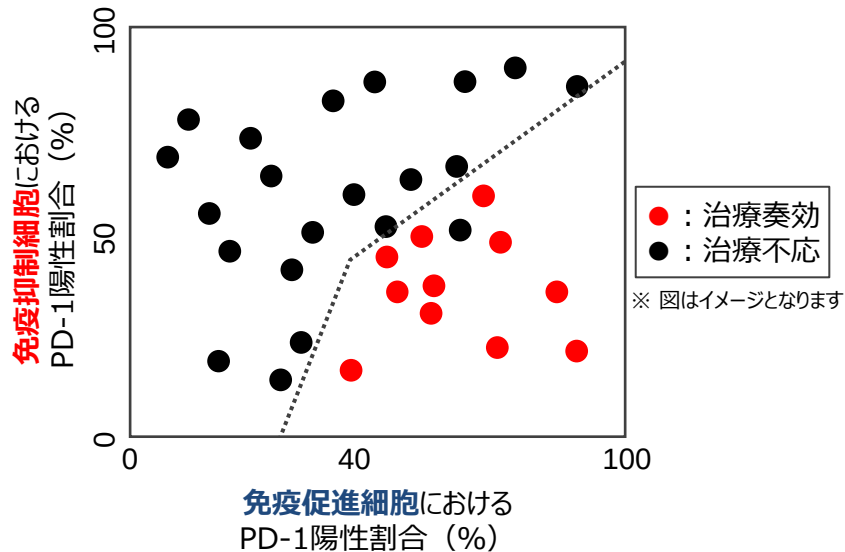


参考：Kumagai, S. et al. The PD-1 expression balance between effector and regulatory T cells predicts the clinical efficacy of PD-1 blockade therapies. *Nature Immunology* 21, 1346–1358 (2020).

### 3-1. ヘルスケアデータへの日立独自AI適用事例①

## 日立のAIを活用して、免疫チェックポイント阻害剤の治療効果を予測するバイオマーカーを同定

B3が求めた患者ごとのロジスティック回帰モデル集約し、  
治療効果あり群を識別する層別境界（……）を定義

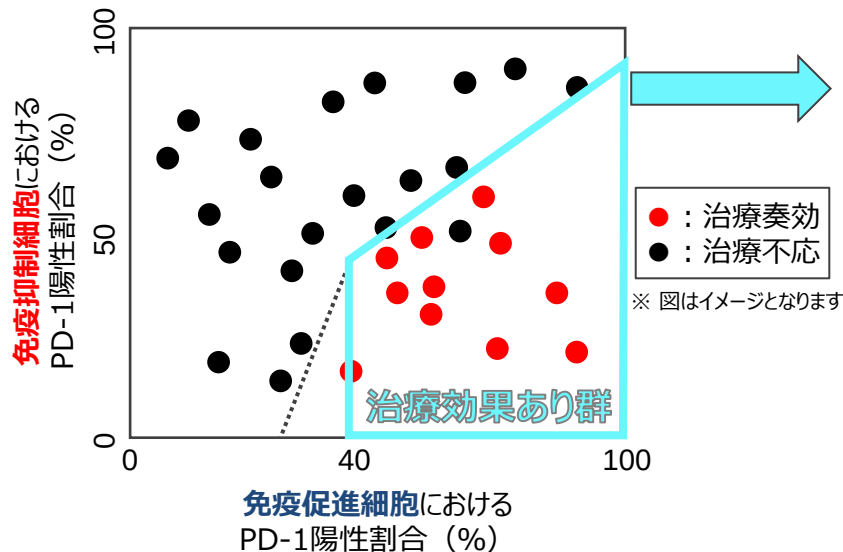


参考 : Kumagai, S. *et al.* The PD-1 expression balance between effector and regulatory T cells predicts the clinical efficacy of PD-1 blockade therapies. *Nature Immunology* 21, 1346–1358 (2020).

### 3-1. ヘルスケアデータへの日立独自AI適用事例①

## 日立のAIを活用して、免疫チェックポイント阻害剤の治療効果を予測するバイオマーカーを同定

B3が求めた患者ごとのロジスティック回帰モデル集約し、  
治療効果あり群を識別する層別境界（……）を定義



以下の2つの条件を満たすグループを**治療効果あり群**と定義

1. **免疫促進細胞にPD-1が40%以上発現**
2. **免疫抑制細胞より免疫促進細胞優位にPD-1が発現**

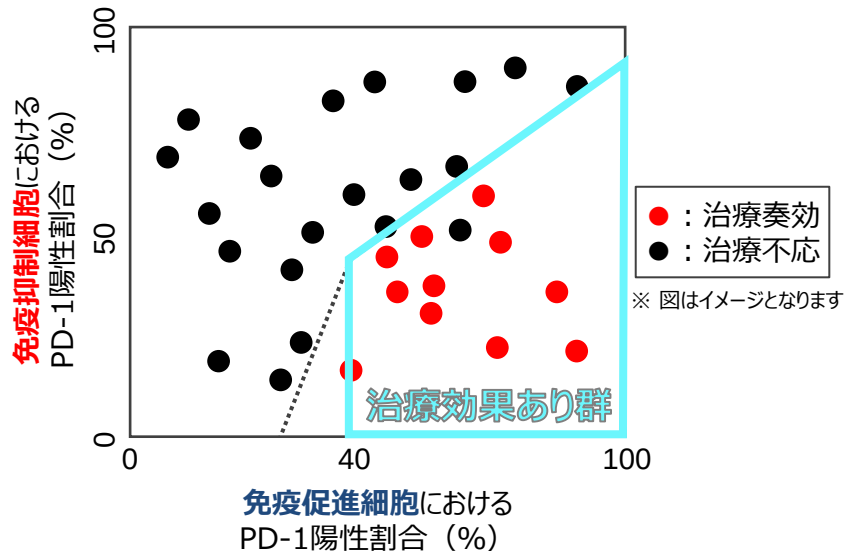
AIの探索結果は、生物学的メカニズムとも合致するか？

参考 : Kumagai, S. et al. The PD-1 expression balance between effector and regulatory T cells predicts the clinical efficacy of PD-1 blockade therapies. *Nature Immunology* 21, 1346–1358 (2020).

### 3-1. ヘルスケアデータへの日立独自AI適用事例①

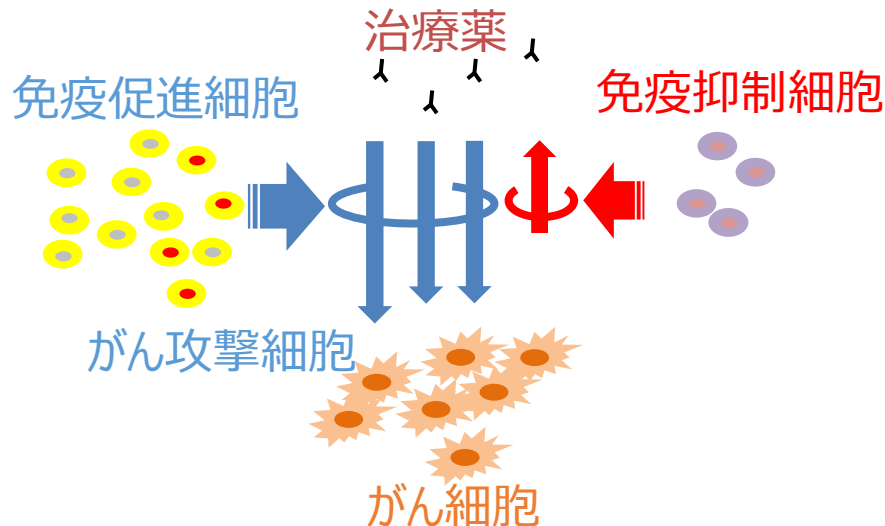
## 日立のAIを活用して、免疫チェックポイント阻害剤の治療効果を予測するバイオマーカーを同定

B3が求めた患者ごとのロジスティック回帰モデル集約し、  
治療効果あり群を識別する層別境界（……）を定義



動物実験モデルに整合（結果の妥当性を強化）

免疫促進細胞と免疫抑制細胞のバランスが治療効果と関連  
[PD-1陽性CD8陽性T細胞数 > PD-1陽性制御性T細胞数]



参考 : Kumagai, S. et al. The PD-1 expression balance between effector and regulatory T cells predicts the clinical efficacy of PD-1 blockade therapies. *Nature Immunology* 21, 1346–1358 (2020).

### 課題：糖尿病性腎症の高性能な予後予測

#### 糖尿病性腎症

- 糖尿病の合併症であり、腎臓中の糸球体のフィルターに穴が開いてしまい、体に戻す必要のあるタンパク質が尿に漏れることで、徐々に腎機能が低下する病気
- 急速に腎機能が低下する患者（Rapid Decliner）が一定数おり、Rapid Declinerへは早期の介入が必要不可欠
  - どのような患者がRapid Declinerとなるのかは不明確

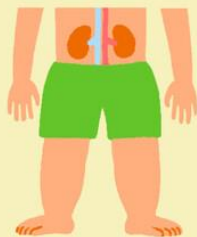
#### 【糖尿病の三大合併症】

##### 糖尿病性神経障害



立ちくらみや手足のしびれ  
などの症状が出現  
足の壊死につながることも

##### 糖尿病性腎症

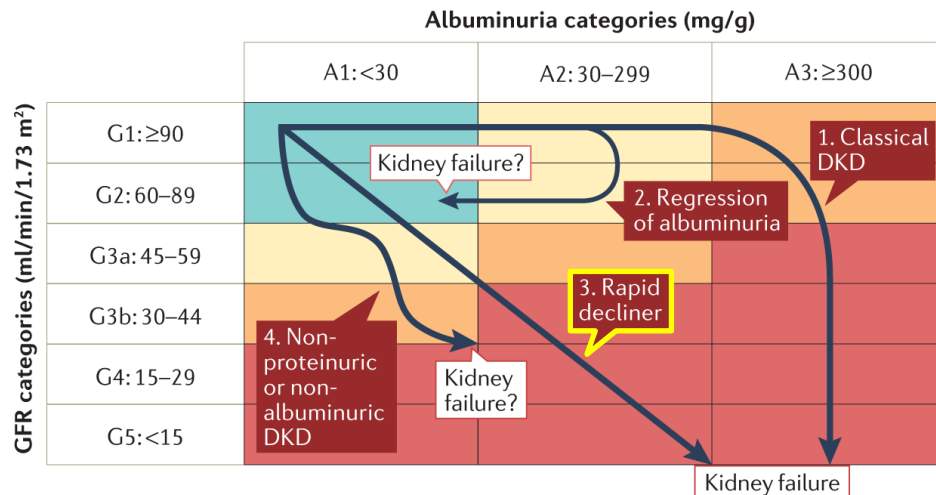


腎臓の機能が低下  
血液透析の原因第一位

##### 糖尿病性網膜症



自覚したときには  
失明寸前のことも

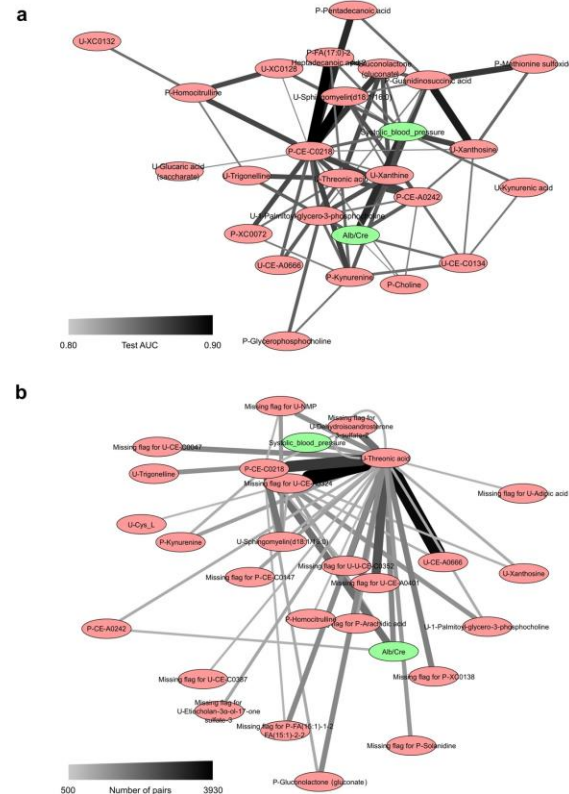
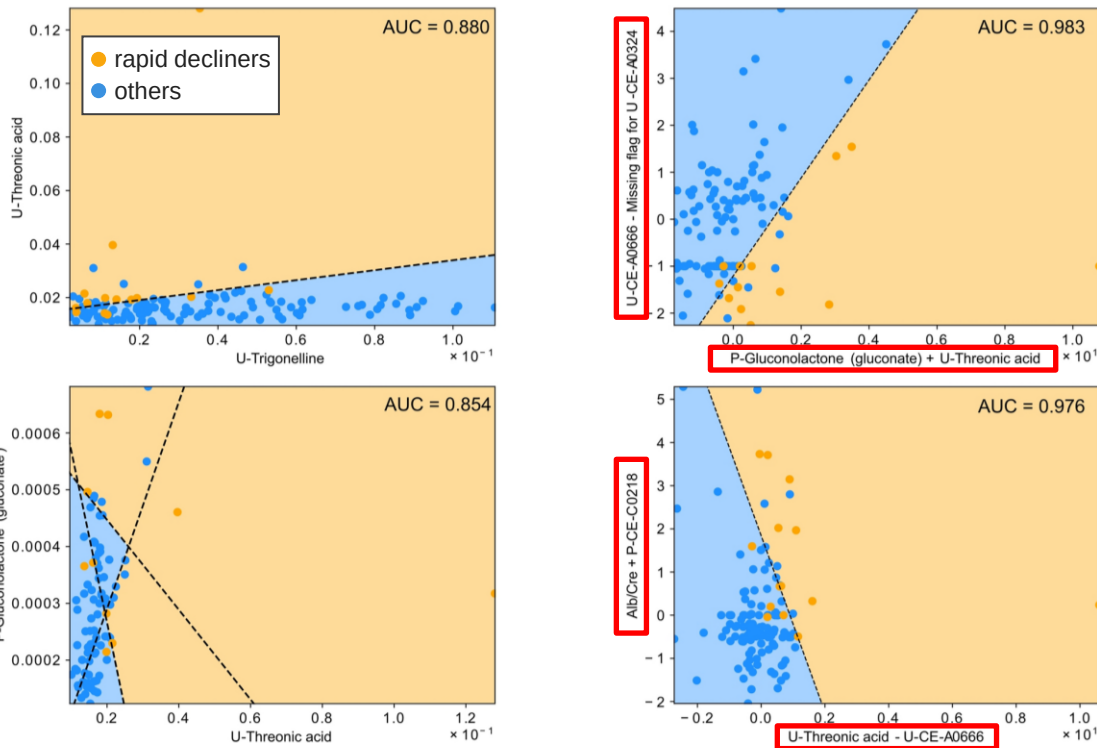


【特定疾病】糖尿病性神経障害、糖尿病性腎症及び糖尿病性網膜症  
（糖尿病の合併症） - LIFULL 介護(ライフ介護) (homes.co.jp)

Trajectories of kidney function in diabetes: a clinicopathological  
update | Nature Reviews Nephrology

© Hitachi, Ltd. 2023. All rights reserved.

- 糖尿病性腎症における予後(30カ月後)を識別可能なマーカーを、メタボロームと臨床情報から探索
- AIが特定代謝物の組み合わせが予後予測に重要であることを示唆

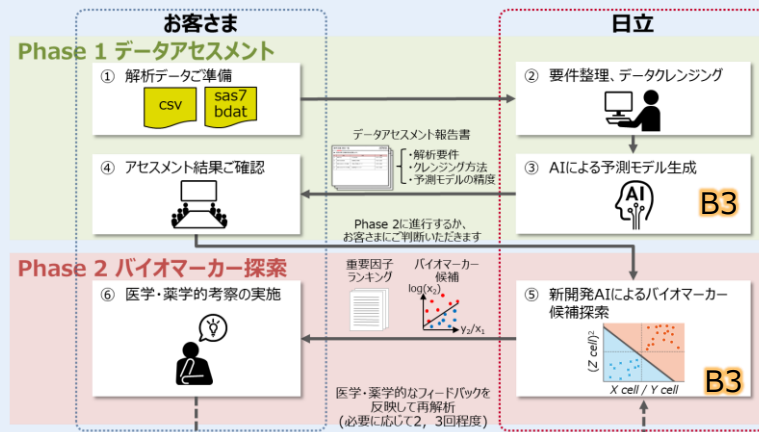


Hirakawa, Y., Yoshioka, K., Kojima, K. *et al.* Potential progression biomarkers of diabetic kidney disease determined using comprehensive machine learning analysis of non-targeted metabolomics. *Sci Rep* **12**, 16287 (2022).

受託解析としてご提供してきたサービスに加え、お客さま自身で「B3」を使い  
解析が行える**クラウド利用型の解析基盤提供サービス**をリリース

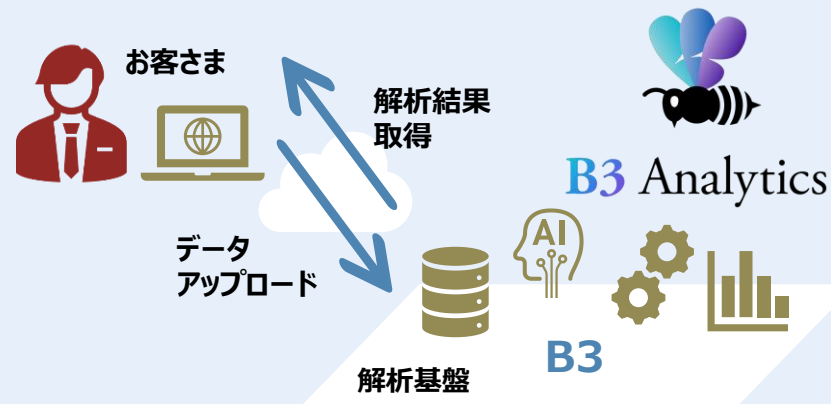
### 今までご提供してきた… **受託型**の解析支援サービス

- 解析要件の整理からご支援
- 日立のデータサイエンティストが、ご要望に合わせ解析や結果可視化等を柔軟に対応
- お客さまが求める支援レベルに応じた、サービスのカスタマイズが可能（個別見積）



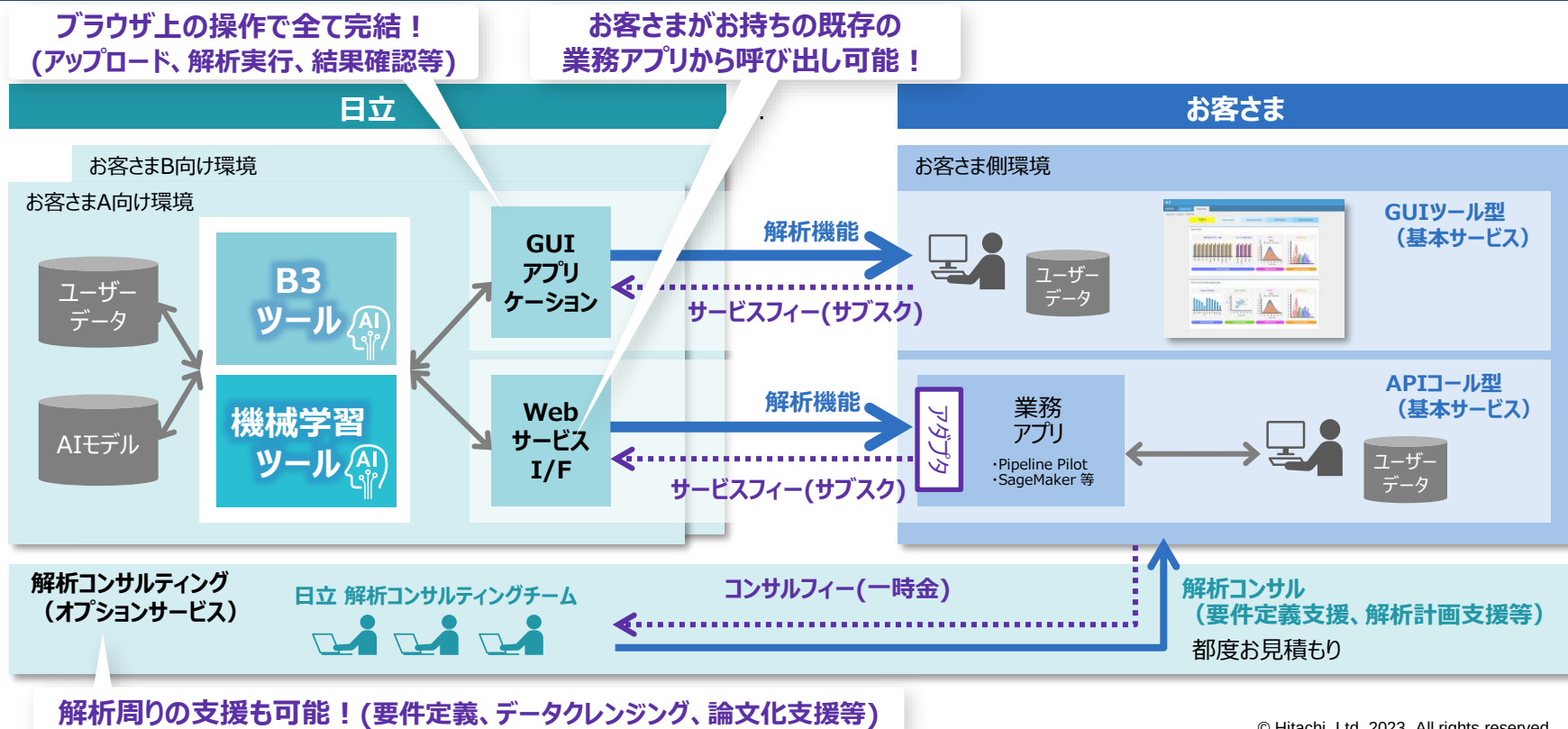
### 今回リリースした… **クラウド利用型**の解析基盤提供サービス

- お客さま自身で、B3を搭載した解析環境を操作
- お客さま自身で、解析実行～結果確認を繰り返し、試行錯誤しながら考察・検証が行える
- 使いたい時に使いたい期間だけ使える（月額利用料）



## 4-1. 今後の展望

Webブラウザ経由で操作できる**GUI**と、外部から呼び出し可能な**API**を基本サービスとし、  
日立のデータサイエンティストによる**解析コンサルティングサービス**もオプションでご提供



### 1 ヘルスケアデータ解析に使える“AI”とは？

- 医師による意思決定には“説明可能性”が求められる
- ヘルスケアデータ解析では“ロジスティック回帰”の説明方法がデファクト

### 2 日立独自説明可能AI

- ロジスティック回帰よりも高い予測性能をもち、かつヘルスケアデータ解析でデファクトとして使われる“ロジスティック回帰”と同じような解釈が可能なAIを開発
  1. B3：患者ごとにカスタムメイドされたロジスティック回帰モデルを有するAI
  2. ADA：ロジスティック回帰でも性能が高くなる因子を生成するAI
- これらのAIを治療効果予測マーカーや予後予測マーカーの探索研究に適用

### 3 今後の展望

- 日立独自説明可能AIをお客さま自身で使っていただけるクラウド利用型のサービスを展開

A woman with her back to the camera, wearing a bright yellow long-sleeved shirt and dark jeans, stands on a green hill. She is looking out over a city skyline on the right and several wind turbines on the left. The entire scene is overlaid with a blue network of lines and glowing nodes, symbolizing connectivity and technology. The sky is a clear, vibrant blue.

Hitachi Social Innovation is  
**POWERING GOOD**