

Predictive Approaches for Low-cost Preventive Medicine Program in Developing Countries

(accepted to KDD'15 Industry and Government track)

Yukino Baba, Hisashi Kashima, Yasunobu Nohara, Eiko Kai, Partha Ghosh, Rafiqul Islam, Ashir Ahmed, Masahiro Kuroda, Sozo Inoue, Tatsuo Hiramatsu, Michio Kimura, Shuji Shimizu, Kunihisa Kobayashi, Koji Tsuda, Masashi Sugiyama, Mathieu Blondel, Naonori Ueda, Masaru Kitsuregawa, Naoki Nakashima

概要

機械学習を応用して予防医療コストを削減、
発展途上国での予防医療普及につながると期待

データ

バングラデシュで収集した
1万5千人の健診データ

結果1

健診コスト削減
1%の高リスク者の見逃しで健診費用を50%削減

結果2

経過観察コスト削減
40%のコストで67%の高リスク者をカバー

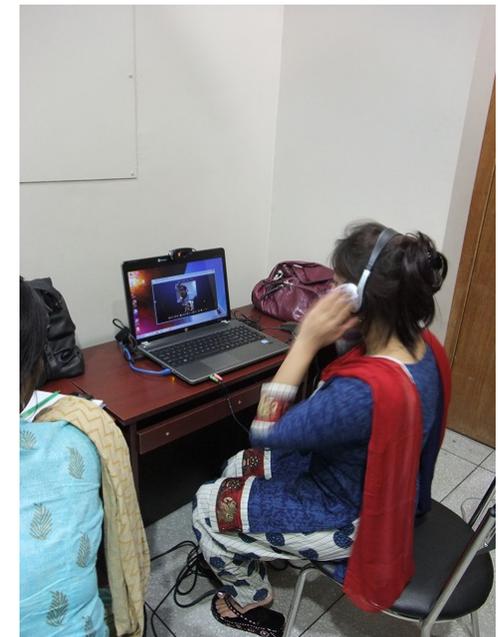
**発展途上国でも生活習慣病が増加、
予防医療が重要だが途上国では費用・人員が不足**

- 先進国特有と考えられていた**生活習慣病**が**発展途上国**でも増加
 - 生活習慣病：生活習慣改善で予防可能な疾患（心血管疾患、がん、糖尿病等）
- 生活習慣病を防ぐには予防医療が重要
- 発展途上国では予防医療に割く**費用・人員が不足**
→ 予防医療の普及が困難

我々の取り組み：Portable Health Clinic

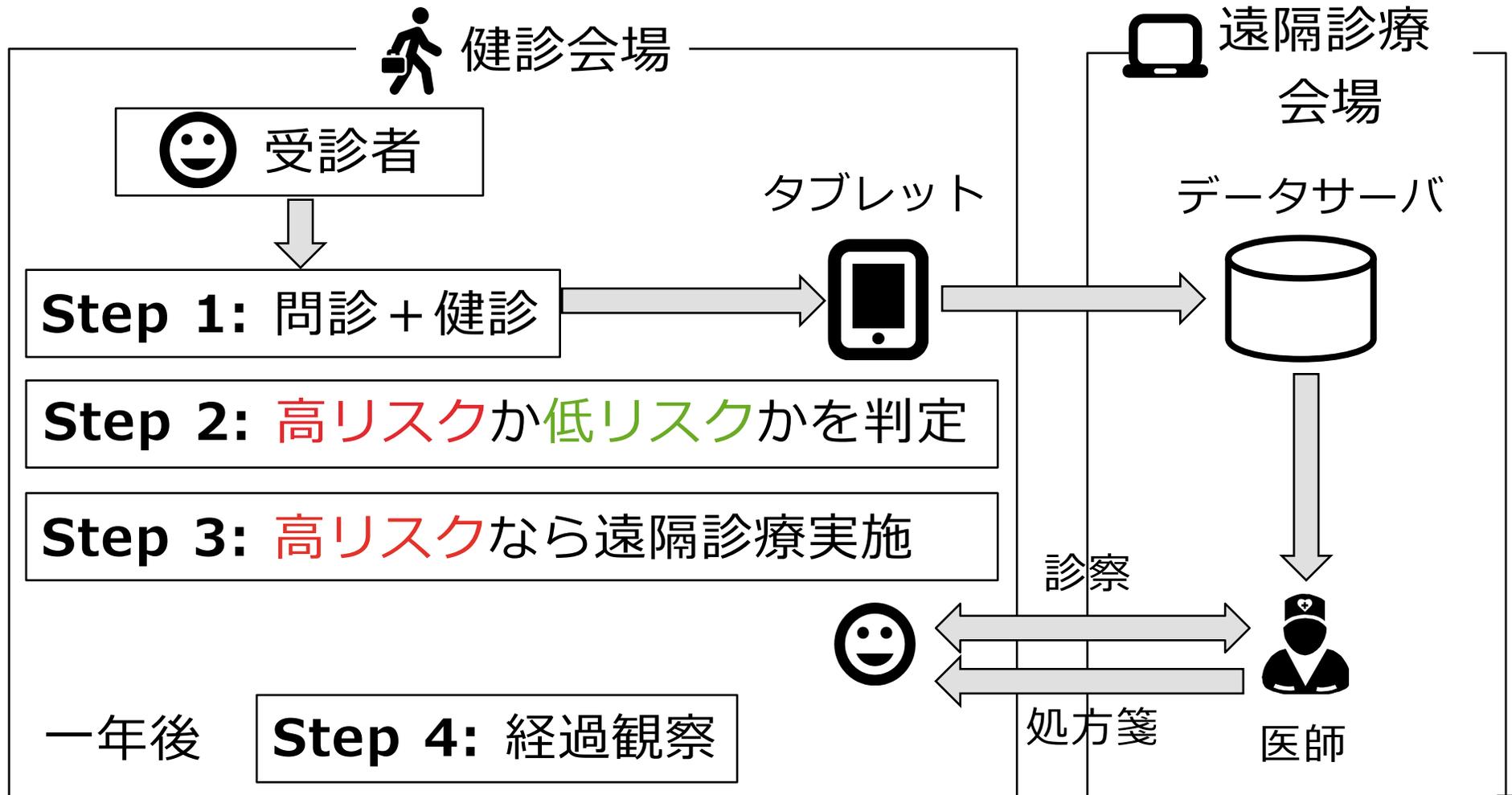
簡易検査・遠隔診療システムを構築、 途上国での予防医療を支援

- 簡易検査：持ち運び容易な検査キット・デバイス
- 遠隔診療：Skypeを通じて医師の診察・薬処方

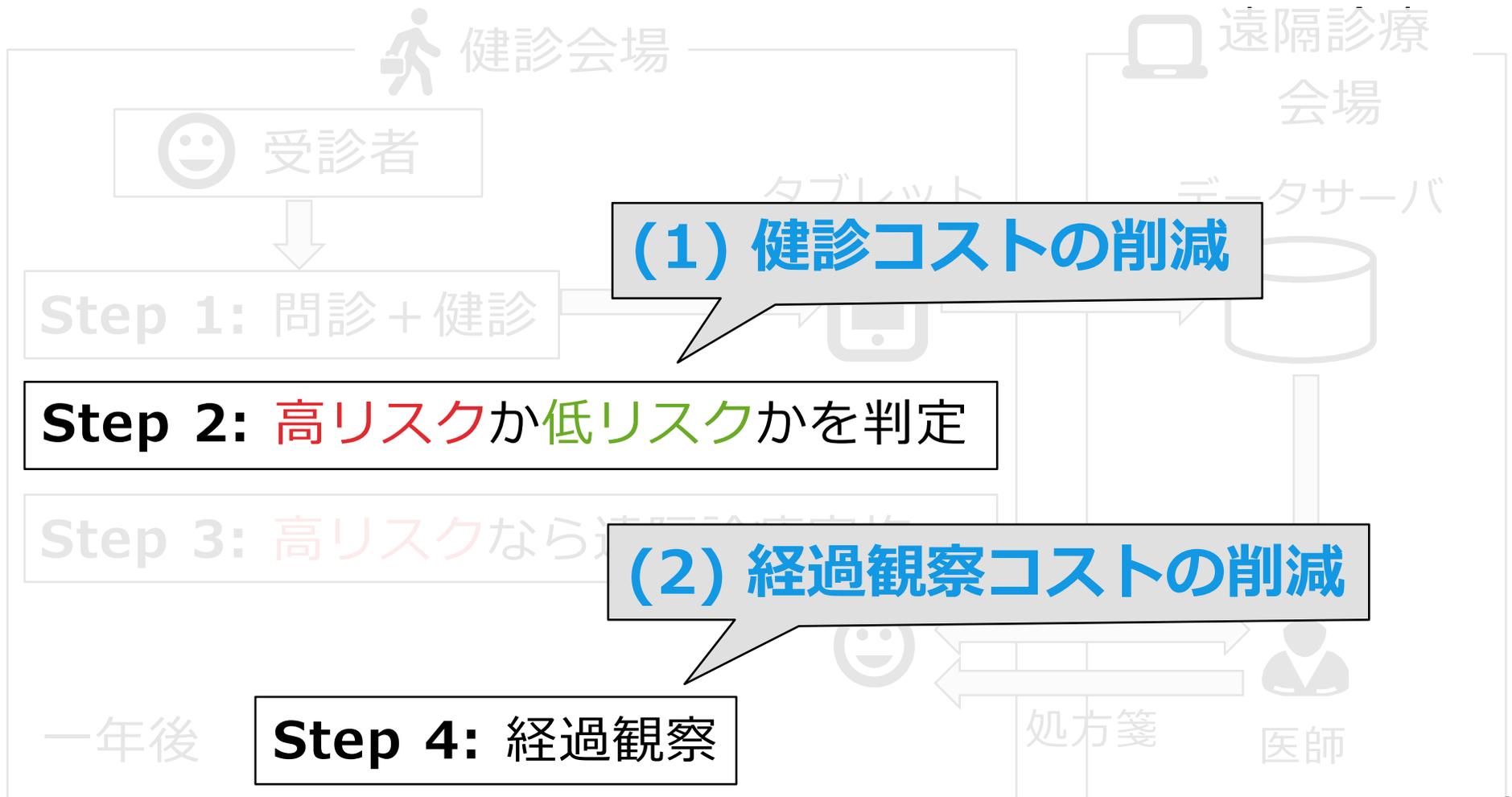


我々の取り組み : Portable Health Clinic

簡易検査により高リスク者を発見、遠隔診療を実施



収集データを用い、機械学習によるコスト削減



(1) 健診コスト削減

健診コスト削減の背景：健診の目的

健診の目的は健康リスクの判定

- 健康リスクの定義：
 - 測定値にもとづき各項目に**Healthy**, **Caution**, **Affected**, **Emergency**のラベルを付与
 - 12項目中一つでも**Affected**以上だと**高リスク**、それ以外は**低リスク**

高リスクの例

ウェスト周囲径	72cm	尿蛋白	±
ウェスト・ヒップ比	0.8	尿糖	+
BMI	24	尿ウロビリノーゲン	±
血圧	135mmHg	不整脈	+
血糖	110mg/dl	体温	36.5℃
脈拍	105	SpO2	94%

健診コスト削減の背景：有料健診

有料健診を一部省略して健診費用を削減したい

- 消耗品代が掛かる健診項目がある（有料健診）
- 一部の有料健診を省略し残りから**健康リスクを予測**できれば健診費用を削減できる

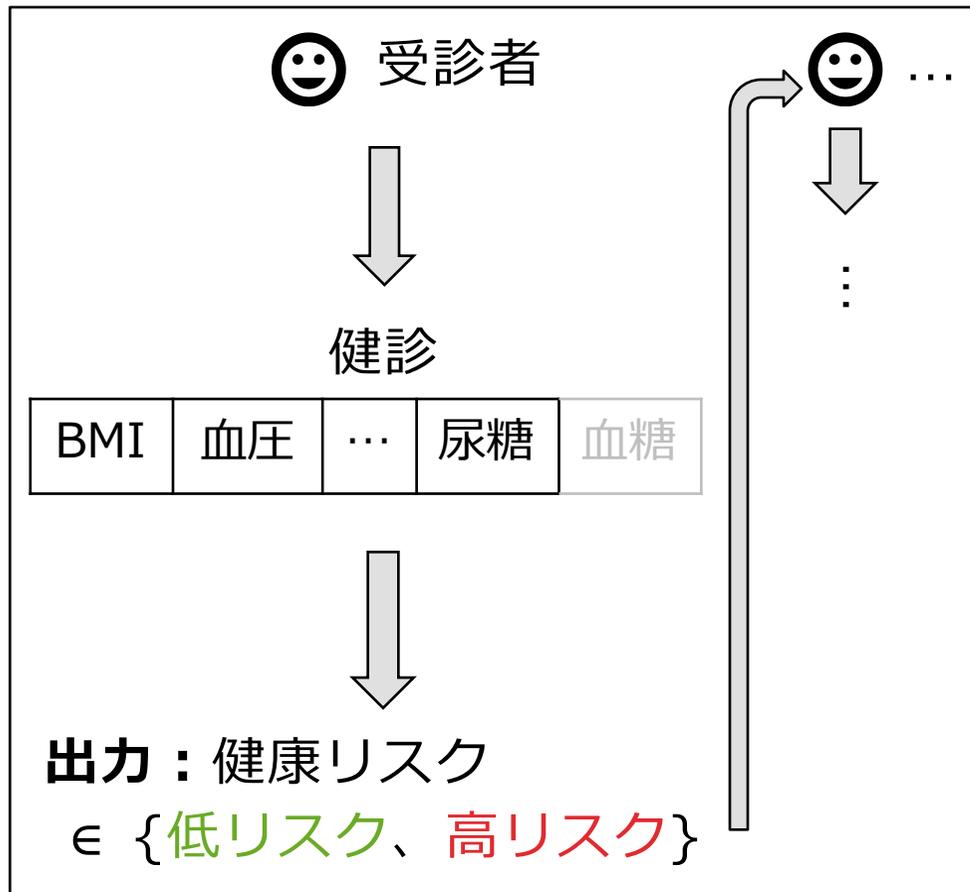
健診項目	費用 (消耗品代)
ウェスト周囲径	0円
ウェスト・ヒップ比	0円
BMI	0円
血圧	0円
血糖	57円
脈拍	0円

健診項目	費用 (消耗品代)
尿蛋白	19円
尿糖	
尿ウロビリノーゲン	
不整脈	0円
体温	0円
SpO2	0円

健康リスク予測問題

ゴール：健診費用を抑えて全員の健康リスクを予測

健診の流れ

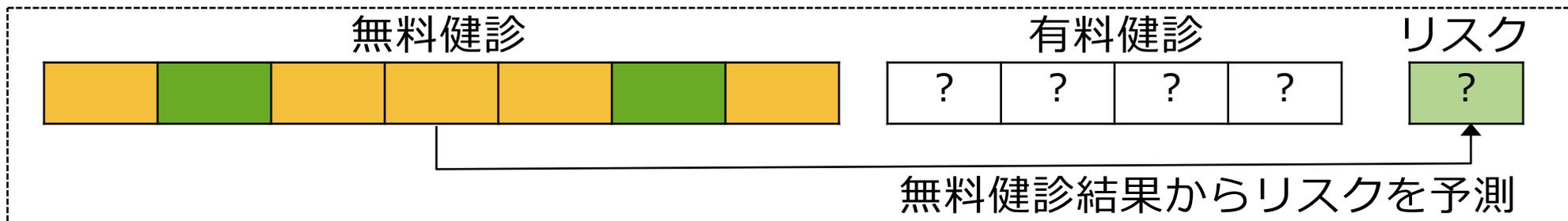


- 健診項目実施順は全員同じ
- 何日目まで実施するかは受診者ごとに制御可能
- 高リスクと予測された受診者には、遠隔診療のために全項目を受診させる
- **ゴール**：健診費用を抑え、全受診者の健康リスクを高精度予測

健康リスク予測手法：予測を一度だけ実施

有料健診を全て省略、無料健診結果から予測

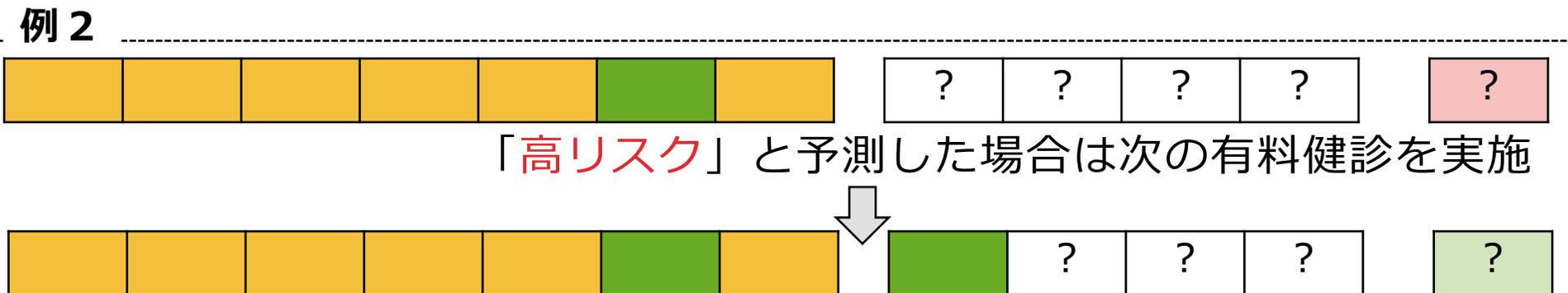
- 最初のN人には無料・有料健診を全て実施、結果を使って予測器を学習
- 残りの受診者には無料健診だけを実施、無料健診結果から健康リスクを予測



健康リスク予測手法：予測を複数回実施

段階的に健診を実施、費用を掛けるべきか慎重に判断

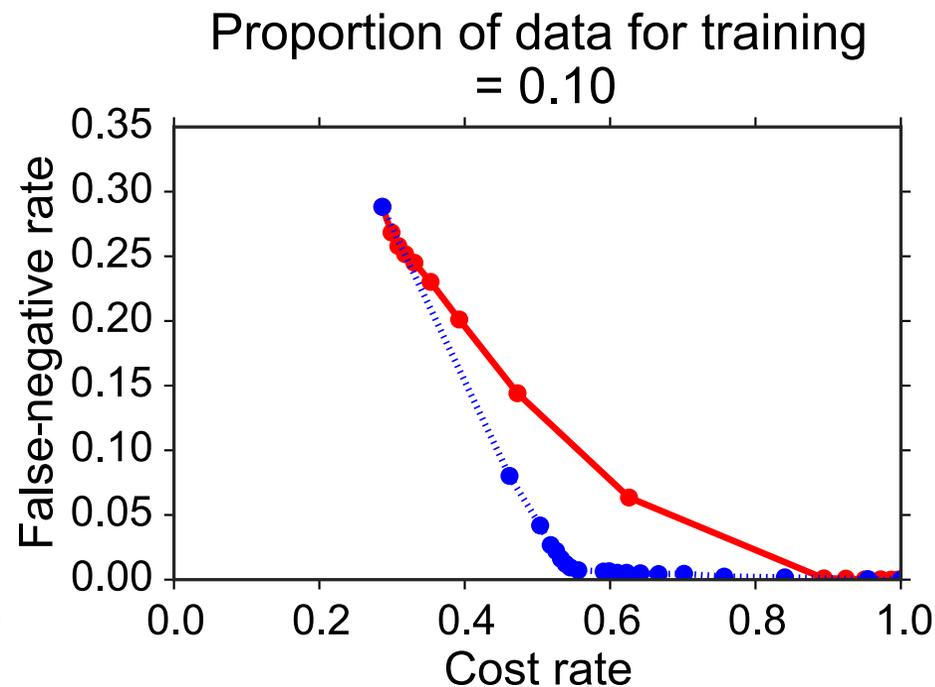
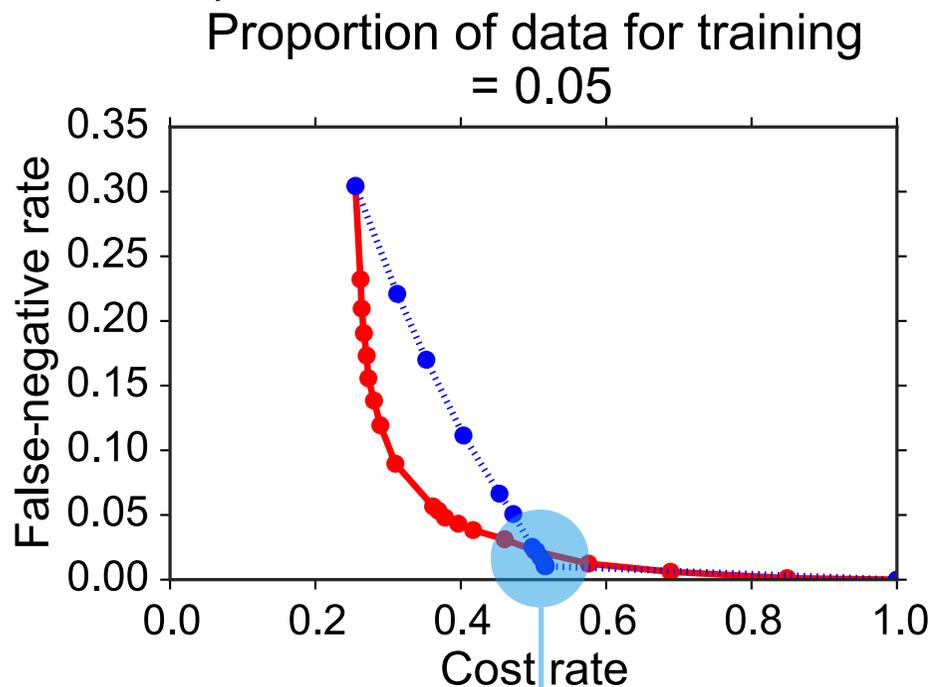
- 有料健診を段階的に実施
- ある時点で「低リスク」と予測されたら、費用を掛ける必要なしと判断し以降を省略



健診コスト削減 実験結果

1%の高リスク者の見逃しで健診費用を50%削減

- Single classifier
- Multiple-classifier



1.1%の見逃しで
全員が全健診を受けた場合の51.5%の健診費用に抑えた

健診コスト削減 結果分析

問診結果が健康リスク予測に有用

- 問診での薬服用履歴情報が健康リスク予測に有用
 - 糖尿病の薬を服用している人は
高リスクの確率が高い
- BMI、血圧、不整脈がAffectedかEmergency :
高リスクの確率が高い
- BMI、血圧、体温がHealthy :
低リスクの確率が高い

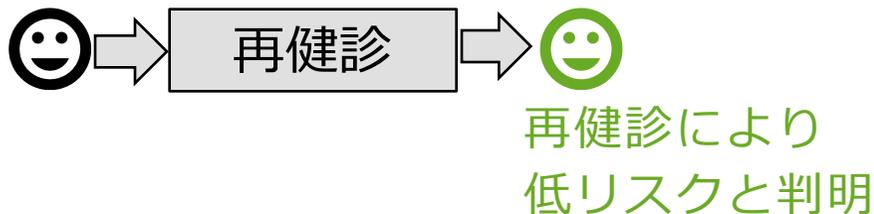
(2) 経過観察コスト削減

経過コスト削減の背景

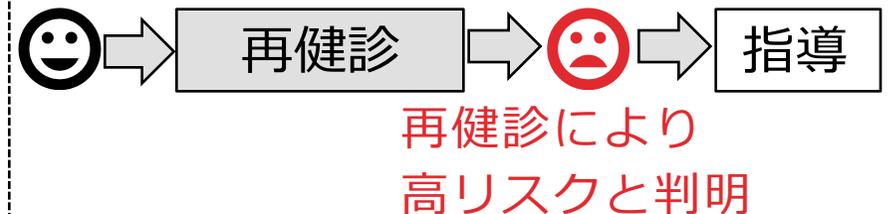
2年め「高リスク」の人の再健診を優先したい

- 経過観察：翌年再健診を実施、**高リスク**者に指導
- 経過観察コスト：再健診に掛かる費用
- 経過観察コスト削減のため、
2年めの**高リスク**者の再健診を優先したい
→1年めの健診結果から**2年めの健康リスクを予測**

再健診を省略したいケース



再健診を優先したいケース



再健診実施順決定問題

ゴール：予算制約内で高リスク者を多く発見

- **入力：**
全員の1年めの健診結果（全項目）、予算制約
- **ゴール：**
予算制約内で、2年め高リスクの人を
できるだけ多く発見できる**再健診実施順**の決定
- 実施順は逐次的に決めて良い
- 再健診費用は受診者に依らず一定

実施順決定手法：「活用のみ」戦略

高リスクである確率が大い受診者を優先

学習データ収集：ランダムに選んだN人に再健診実施

 予測器を学習

活用：
高リスク確率最大の受診者*i**を選択
$$i^* = \underset{i}{\operatorname{arg\,max}} p(y_i = 1 \mid \mathbf{x}_i)$$

受診者*i*が
高リスクである確率

 再健診実施、
結果を使って予測器を更新

予算に達したら終了

実施順決定手法：「探索と活用」戦略

分類器学習に寄与する受診者も選択

学習データ収集：ランダムに選んだN人に再健診実施

 予測器を学習

確率 ϵ

確率 $1 - \epsilon$

探索：

確信度最低の受診者 i^* を選択

$$i^* = \arg \min_i |p(y_i = 1 | \mathbf{x}_i) - 0.5|$$

活用：

高リスク確率最大の受診者 i^* を選択

$$i^* = \arg \max_i p(y_i = 1 | \mathbf{x}_i)$$

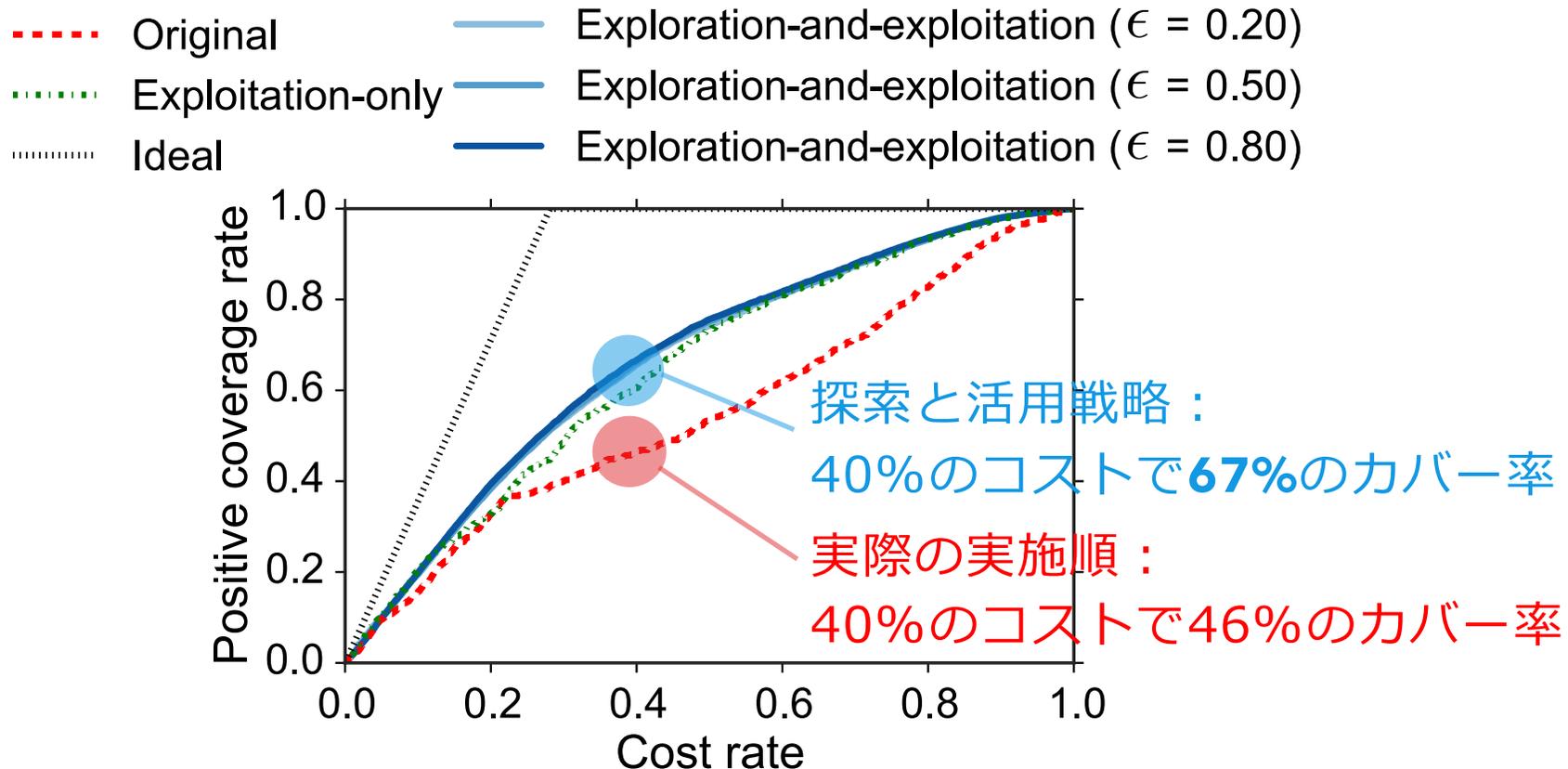


再健診実施、
結果を使って予測器を更新

予算に達したら終了

経過観察コスト削減 実験結果

40%のコストで67%の高リスク者をカバー



- 「探索」追加によるカバー率向上を確認
- ϵ の値による影響は小さい

経過観察コスト削減 結果分析

問診結果が翌年の健康リスク予測にも有用

- 定期的に医師の診察を受けていない人：
翌年**高リスク**である確率が高い
- 歩行スピードが早い人は：
翌年**低リスク**である確率が高い

まとめ

機械学習を応用して予防医療コストを削減、
発展途上国での予防医療普及につながると期待

データ

バングラデッシュで収集した
1万5千人の健診データ

結果1

健診コスト削減
1%の高リスク者の見逃しで健診費用を50%削減

結果2

経過観察コスト削減
40%のコストで67%の高リスク者をカバー