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

## Highlights

Predictive modeling supports low-cost preventive medicine in Bangladesh

<table>
<thead>
<tr>
<th>Data</th>
<th>Health checkup results of 15K subjects in Bangladesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result #1</td>
<td>Health checkup cost reduction: 50% cost reduction with 1% of false negative results</td>
</tr>
<tr>
<td>Result #2</td>
<td>Follow-up checkup cost reduction: 67% coverage of high-risk subjects with 40% of the total budget</td>
</tr>
</tbody>
</table>
Our project: Portable Heath Clinic (PHC)

PHC is an efficient preventive medicine service

- **Portable package** for health checkups
- **Tele-consultation** service using Skype
Our project: Portable Health Clinic (PHC)

2-year-long PHC field study in Bangladesh

Step 1: Checkup

Step 2: Risk assignment (low or high)

Step 3: Tele-consultation for high-risk subject

One year later

Step 4: Follow-up

* Doctor image is designed by Freepik.com
Scope of this study

ML methods for cost-efficient PHC service

1. SUBJECT RISK PREDICTION FOR CHECKUP COST REDUCTION
   - Step 1: Checkup
   - Step 2: Risk assignment (low or high)

2. FUTURE RISK PREDICTION FOR EFFICIENT FOLLOW-UP PROCESS
   - Step 4: Follow-up

* Doctor image is designed by Freepik.com
1 SUBJECT RISK PREDICTION FOR CHECKUP COST REDUCTION
Background: Health checkup

Health checkup determines risk-level of subject

For each subject, our criterion assigns Healthy, Caution, Affected, or Emergency to each of 12 checkup items

<table>
<thead>
<tr>
<th>Item</th>
<th>Value</th>
<th>Item</th>
<th>Value</th>
<th>Item</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waist</td>
<td>72cm</td>
<td>Pulse rate</td>
<td>105</td>
<td>Urine-protein</td>
<td>±</td>
</tr>
<tr>
<td>Waist/hip ratio</td>
<td>0.8</td>
<td>Arrhythmia</td>
<td>+</td>
<td>Urine-sugar</td>
<td>+</td>
</tr>
<tr>
<td>BMI</td>
<td>24</td>
<td>Body temperature</td>
<td>36.5℃</td>
<td>Urine-urobilinogen</td>
<td>±</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>135mmHg</td>
<td>Oxygen saturation</td>
<td>94%</td>
<td>Blood sugar</td>
<td>110mg/dl</td>
</tr>
</tbody>
</table>

😊 Low-risk
If a subject is assigned with Healthy or Caution in all the items

😢 High-risk
If a subject is assigned with Affected or Emergency in at least one item
Background: Cost of health checkup

High-cost tests make health checkup expensive

Consumables cost per checkup item (in USD)

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
<th>Item</th>
<th>Cost</th>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waist</td>
<td>0</td>
<td>Pulse rate</td>
<td>0</td>
<td>Urine-protein</td>
<td></td>
</tr>
<tr>
<td>Waist/hip ratio</td>
<td>0</td>
<td>Arrhythmia</td>
<td>0</td>
<td>Urine-sugar</td>
<td>0.19</td>
</tr>
<tr>
<td>BMI</td>
<td>0</td>
<td>Body temperature</td>
<td>0</td>
<td>Urine-urobilinogen</td>
<td></td>
</tr>
<tr>
<td>Blood pressure</td>
<td>0</td>
<td>Oxygen saturation</td>
<td>0</td>
<td>Blood sugar</td>
<td>0.57</td>
</tr>
</tbody>
</table>

We can save the cost if we skip high-cost tests and predict the risk-level from the rest of tests.
Subject risk prediction problem

Goal: To predict risk-level for all subjects with low checkup costs

Checkup process

We can control when to stop checkup for each subject

Subjects who are predicted as high-risk are required to take all the tests
Single-classifier strategy

- We offer all the tests to the first N subjects, obtain groundtruth, and train a classifier.
- The rest of the subjects are only offered the low-cost tests.

Risk-level is predicted from the results of low-cost tests.
Multiple-classifiers strategy

Multiple-classifiers strategy carefully selects subjects who need the high-cost tests

If a classifier predicts a subject as high-risk, the next high-cost test will be conducted.

Case 1
Low-cost tests: Process is terminated when the subject is predicted as low-risk.

Case 2
Next high-cost test is offered when the subject is predicted as high-risk.
Results of checkup cost reduction

50% cost reduction with 1% of false negative

Multiple classifier strategy achieved false-negative rate=1.1% when cost rate=51.5%
FUTURE RISK PREDICTION FOR EFFICIENT FOLLOW-UP PROCESS
Follow-up process takes care of subjects who are at high-risk in the 2nd year.

If a budget is limited, we should set the priority to subjects.

We determine the priority by predicting the risk-level of the 2nd year.
Follow-up priority detection problem

Goal: To determine the follow-up priority under the budget constraint

- **Input:**
  Checkup results of the 1st year, Budget constraint

- **Output:**
  Priority of subjects to examine as many of the high-risk subjects as possible

<table>
<thead>
<tr>
<th>Subject</th>
<th>😊</th>
<th>😊</th>
<th>😊</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of 2nd year</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Results of 1st year

<table>
<thead>
<tr>
<th>Subject</th>
<th>😒</th>
<th>😒</th>
<th>😊</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of 2nd year</td>
<td>😒</td>
<td>😒</td>
<td>😊</td>
</tr>
</tbody>
</table>

Priority

high 🟢 low
Exploitation-only strategy

Exploitation-only strategy puts priority to subjects who are likely to be at high-risk

Initial classifier training with N samples

Exploitation:
Select the subject who is the most likely to be at high-risk

\[ i^* = \arg \max_i p(y_i = 1 \mid x_i) \]

Providing the checkup to the selected subject,
Updating the classifier with the output

Results of 1st year
**Exploration-and-exploitation strategy**

** Exploration-and-exploitation strategy puts priority to informative subjects**

- **Initial classifier training** with N samples

  - Prob. of $\epsilon$
  - Prob. of $1 - \epsilon$

  - **Exploration**:
    - Select the subject with the **least certainty of being at high-risk**
    - $i^* = \arg\min_i |p(y_i = 1 | x_i) - 0.5|$

  - **Exploitation**:
    - Select the subject who is the **most likely to be at high-risk**
    - $i^* = \arg\max_i p(y_i = 1 | x_i)$

- Providing the checkup to the selected subject,
- **Updating the classifier** with the output
Results of checkup cost reduction

67% coverage of high-risk subjects with 40% of the total budget

- Exploration-and-exploitation ($\epsilon = 0.20$)
- Exploration-and-exploitation ($\epsilon = 0.50$)
- Exploration-and-exploitation ($\epsilon = 0.80$)

Original procedure

- Coverage = 46% when cost rate = 40%

Exploration-and-exploitation strategy

- Coverage = 67% when cost rate = 40%
## Summary

Predictive modeling supports low-cost preventive medicine in Bangladesh

<table>
<thead>
<tr>
<th>Data</th>
<th>Health checkup results of 15K subjects in Bangladesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result #1</td>
<td>Health checkup cost reduction: 50% cost reduction with 1% of false negative results</td>
</tr>
<tr>
<td>Result #2</td>
<td>Follow-up checkup cost reduction: 67% coverage of high-risk subjects with 40% of the total budget</td>
</tr>
</tbody>
</table>