Abstract: P571

Accuracy of a machine learning program to correctly triage incoming SMS text replies from a successful cardiovascular SMS-based secondary prevention program

Authors:
N Lowres¹, A Duckworth², CK Chow³, A Thiagalingam⁴, J Redfern³, ¹Heart Research Institute and University of Sydney - Camperdown - Australia, ²Sandstone Technology - Sydney - Australia, ³Westmead Applied Research Centre, University of Sydney - Sydney - Australia, ⁴Westmead Hospital - Westmead - Australia

Topic(s):
m-Health

Citation:
National Heart Foundation Vanguard Grant; National Health and Medical Research Council Project Grant

BACKGROUND: Cardiovascular SMS text programs are effective alternate secondary prevention programs for cardiac risk factor reduction and can be delivered as one-way or two-way communication. However, people text back regularly, leading to staffing costs to monitor replies. If you could reduce the need for staff review by 60-70%, costs and scalability of text programs would substantially improve.

PURPOSE: To develop and assess accuracy of a machine-learning (ML) program to ‘triage’ and identify texts requiring review/action.

METHODS: We manually reviewed and classified all replies received from two ‘TEXT ME’ cardiovascular secondary prevention programs. Simultaneously a ML model was developed to classify texts and determine those needing a reply (figure). Comparison of ML models included ‘Naïve Bayes’, ‘random forest decision trees’, and ‘gradient boosted trees’, along with comparison to ‘convolutional neural network’ and ‘recurrent neural network’ classification approaches. ‘Natural language programming’ was evaluated however this presented challenges in relation to text content due to non-standard English grammar, frequent use of non-standard abbreviations, and spelling errors. The ML program was trained with 70% of the data-set and accuracy was tested with 30%.

RESULTS: Manual review of 3118 text replies revealed that only one text was considered urgent, and only 21% required review/action: categorisation was not straight forward due to complexity of texts often containing more than one sentiment (table). The ML program was able to correctly classify 84% of texts into the designated 12 categories. The sensitivity for correctly identifying the need for health professional review was 94% (6.4% false negatives; 3.6% false positives); but with addition of ‘heuristics’ (e.g. searching for specific keywords, question marks etc) sensitivity increased to 97% (2.9% false negatives; 7.3% false positives). Therefore, health professionals would only have to review 27% (true + false positives) of all text replies.

CONCLUSIONS: The ML program has high sensitivity identifying text replies requiring health professional input and a low false negative rate indicating few messages needing response would be missed. Thus, introduction of the program could significantly reduce the workload of health professionals, leading to substantial improvements in scalability and capacity of text-based programs. The future implications for this technology are vast, including utilisation in other interactive mHealth interfaces and cardiovascular health ‘apps’.

<table>
<thead>
<tr>
<th>REVIEW REQUIRED</th>
<th>Health Question/concern (13%)</th>
<th>Admin request (4.5%)</th>
<th>Request to STOP (3%)</th>
<th>Ceased smoking (0.8%)</th>
<th>SMS not delivered (0.4%)</th>
<th>Urgent/distress (0.03%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO REVIEW REQUIRED</td>
<td>General statement (33%)</td>
<td>Statement of thanks (23%)</td>
<td>Reporting good health (11%)</td>
<td>Blank message (6%)</td>
<td>Unrelated/accidental (4%)</td>
<td>Emoticon only (2.4%)</td>
</tr>
</tbody>
</table>
Abstract:

P571

Accuracy of a machine learning program to correctly triage incoming SMS text replies from a successful cardiovascular SMS-based secondary prevention program

Authors:
N Lowres1, A Duckworth2, CK Chow3, A Thiagalingam4, J Redfern3
1 Heart Research Institute and University of Sydney - Camperdown - Australia,
2 Sandstone Technology - Sydney - Australia,
3 Westmead Applied Research Centre, University of Sydney - Sydney - Australia,
4 Westmead Hospital - Westmead - Australia,

Topic(s): mHealth

Citation:

BACKGROUND: Cardiovascular SMS text programs are effective alternate secondary prevention programs for cardiac risk factor reduction and can be delivered as one-way or two-way communication. However, people text back regularly, leading to staffing costs to monitor replies. If you could reduce the need for staff review by 60-70%, costs and scalability of text programs would substantially improve.

PURPOSE: To develop and assess accuracy of a machine-learning (ML) program to 'triage' and identify texts requiring review/action.

METHODS: We manually reviewed and classified all replies received from two 'TEXT ME' cardiovascular secondary prevention programs. Simultaneously a ML model was developed to classify texts and determine those needing a reply (figure). Comparison of ML models included 'Naïve Bayes', 'random forest decision trees', and 'gradient boosted trees', along with comparison to 'convolutional neural network' and 'recurrent neural network' classification approaches. 'Natural language programming' was evaluated however this presented challenges in relation to text content due to non-standard English grammar, frequent use of non-standard abbreviations, and spelling errors. The ML program was trained with 70% of the data-set and accuracy was tested with 30%.

RESULTS: Manual review of 3118 text replies revealed that only one text was considered urgent, and only 21% required review/action: categorisation was not straightforward due to complexity of texts often containing more than one sentiment (table). The ML program was able to correctly classify 84% of texts into the designated 12 categories. The sensitivity for correctly identifying the need for health professional review was 94% (6.4% false negatives; 3.6% false positives); but with addition of 'heuristics' (e.g. searching for specified keywords, question marks etc) sensitivity increased to 97% (2.9% false negatives; 7.3% false positives). Therefore, health professionals would only have to review 27% (true + false positives) of all text replies.

CONCLUSIONS: The ML program has high sensitivity identifying text replies requiring health professional input and a low false negative rate indicating few messages needing response would be missed. Thus, introduction of the program could significantly reduce the workload of health professionals, leading to substantial improvements in scalability and capacity of text-based programs. The future implications for this technology are vast, including utilisation in other interactive mHealth interfaces and cardiovascular health 'apps'.

REVIEW REQUIRED

Health Question/concern (13%)
1. Admin request (4.5%)
2. Request to STOP (3%)
3. Ceased smoking (0.8%)
4. SMS not delivered (0.4%)
5. Urgent/distress (0.03%)

NO REVIEW REQUIRED

General statement (33%)
1. Statement of thanks (23%)
2. Reporting good health (11%)
3. Blank message (6%)
4. Unrelated/accidental (4%)
5. Emoticon only (2.4%)

Manual TEXT categorisation → Establish machine learning model → Compare and revise models → Final model

Train model with data → Test model performance → Final model

I have just flown back from UK and my blood pressure is really high, 190/105 what should I do to get it down?

I am experiencing a very dry throat, is the from one of the medications...

I would like to thank you for all the support. As it is nearing 12 months I no longer require text meds sms. I am fully recovered and appreciate your help over this period.