Identifying probable PTSD using machine learning

Dan Leightley PhD
King’s Centre for Military Health Research
King’s College London

@_Dr_Daniel | @kcmhr
PTSD: the focus of the media...

Rates of post-trauma stress rise in British military veterans

Kate Kelland

LONDON: Oct 8 (Reuters) - Conflicts in Iraq and Afghanistan may have led to an increase in rates of post-traumatic stress disorder (PTSD), according to new research looking at the mental health of Britain’s soldiers.

PTSD rates increase in UK military personnel, research suggests

Rise in condition mainly seen in veterans who deployed to Iraq and Afghanistan

Higher levels of PTSD among veterans, says study

By Laura Issava
Bbc Health

2 October 2018

Post-traumatic stress disorder of serving Army personnel and military veterans has increased in the last 10 years, a new study suggests.
The prevalence of alcohol misuse has decreased from 2004/06, whereas the prevalence of PTSD has increased from 4% to 6%.

The ultimate goal
Machine learning as an aid... for now

- The risks:
  - **Perception**: Deep Mind and Greenwich hospital
  - **Understanding**: ‘Black box’
  - **Accountability**: ‘Blame’
  - **Human** factor: ‘Loss of jobs’

Work on-going to mitigate these risks
Other work...

- Alcohol misuse (*InDEx*)
  - Personalised messaging and content

- Improving physical activity (*HeadSmart*)
  - User insights and early detection

- Detecting veterans in clinical records
  - To provide further care and support
Approach

Induction of probable PTSD
No indication PTSD

Features contributing to a prediction


**Self-report questionnaire data**
13,690 completed PCL-C
22 variables

**Classical machine learning**
Support Vector Machines
Random Forests
Artificial Neural Network
Bagging

80% retained for training
20% used for testing

**Prediction**
## Approach


<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support Vector Machines</td>
<td>0.91</td>
<td>0.70</td>
<td>0.92</td>
<td>0.74</td>
</tr>
<tr>
<td>Random Forests</td>
<td>0.97</td>
<td>0.60</td>
<td>0.98</td>
<td>0.64</td>
</tr>
<tr>
<td>Artificial Neural Networks</td>
<td>0.89</td>
<td>0.61</td>
<td>0.92</td>
<td>0.45</td>
</tr>
<tr>
<td>Bagging</td>
<td>0.95</td>
<td>0.69</td>
<td>0.96</td>
<td>0.55</td>
</tr>
</tbody>
</table>
### Approach

**Machine learning analysis is helping us to determine WHICH variables are most important**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Rank 1</th>
<th>Rank 2</th>
<th>Rank 3</th>
<th>Rank 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support Vector Machines</td>
<td>AUDIT Score</td>
<td>GHQ-12 score</td>
<td>Age (years)</td>
<td>Consumes alcohol (y/n)</td>
</tr>
<tr>
<td>Random Forests</td>
<td>Gender</td>
<td>AUDIT Score</td>
<td>GHQ-12 score</td>
<td>Service type</td>
</tr>
<tr>
<td>Artificial Neural Networks</td>
<td>GHQ-12 score</td>
<td>AUDIT case (y/n)</td>
<td>AUDIT Score</td>
<td>Consumes alcohol (y/n)</td>
</tr>
<tr>
<td>Bagging</td>
<td>Age</td>
<td>Consumes alcohol (y/n)</td>
<td>Smoking</td>
<td>GHQ-12 score</td>
</tr>
</tbody>
</table>


You can use machine learning too…

Example: https://tinyurl.com/y9xlrll0

1. Publicly available dataset downloaded from UK Data Archive
2. Upload to Microsoft Machine Learning Studio
3. Training ML classifier
4. Define training and testing set
5. Evaluate
Next steps

Risk Factors
Identification of risk factors contributing to an outcome

Real-time support

Cloud-based – recurrent networks

EHR
Self-report
Wearable: FitBit, Apple Watch
Phone data
Thank You

daniel.leightley@kcl.ac.uk