A Study of Automated Sleep Apnoea Detection using Alerte Digital Health’s Artificial Intelligence System

Luke Bollam, Dion Hulshoff, Marcus Turewicz, Razali Mohamad, Jack Philpott

GOALS

• To enable pre-screening for common sleep conditions at home.
• Create a system that utilises a smartphone to provide immediate feedback to patients.
• Develop a non-invasive wearable that still has most of the capabilities of a full PSG system.
• Automatic identification of common sleep respiratory conditions, using AI.

METHODOLOGY

94 patients with varying levels of sleep related respiratory issues were selected from a database. Respiratory events were subsampled to 20 seconds of data each, and an equal number of ‘no-event’ samples were created by randomly pulling 20 second windows between events. Samples included the ECG, EMG, heart rate, Spo2 and Sound signals. 70% of the samples were used for training, the remaining 30% were reserved for evaluation of the trained model.

The pre-processing step involved a range of filtering and down-sampling techniques applied to the signals, in order to reduce the dimensionality of the problem and improve the feasibility of running the AI in real-time on a smartphone.

The neural network structure used was a combination of convolutional layers, bidirectional recurrent layers and feed forward layers. Convolutional layers excel at finding important local features, recurrent layers allow for temporal dependencies to form across the sequence and feed forward layers are useful for reducing dimensionality down to the two labels. The structure was inspired by the way a PSG study is labelled, where a sleep expert will look for features in the signals before, after and during the event, and look for changes in features both forwards and backwards (for example a change in heart rate or an oxygen desaturation a few seconds after an obstruction).

Input Data

The AI model was trained on the training dataset in batches of 50. After each batch, a parameter update was applied to the neural network in a process called batch learning. Every 50 batches, the entire evaluation dataset was pushed through the AI model and an accuracy metric was calculated. The training was considered complete when the accuracy on the evaluation dataset had peaked and then begun to decrease again as the model overfitted to the training dataset.

RESULTS

Four AI models were created to evaluate the effectiveness of the AI in detecting different types of respiratory events. The results can be seen in the table below.

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Training Examples</th>
<th>Evaluation Examples</th>
<th>Accuracy</th>
<th>AUC-ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obstructive Apnoea</td>
<td>2867</td>
<td>1229</td>
<td>84.20%</td>
<td>0.90</td>
</tr>
<tr>
<td>Central Apnoea</td>
<td>1774</td>
<td>760</td>
<td>85.33%</td>
<td>0.91</td>
</tr>
<tr>
<td>Hypopnoea</td>
<td>9342</td>
<td>4004</td>
<td>69.86%</td>
<td>0.70</td>
</tr>
<tr>
<td>All*</td>
<td>14095</td>
<td>6040</td>
<td>74.55%</td>
<td>0.83</td>
</tr>
</tbody>
</table>

* All refers to a combination of Obstructive, Central and Mixed Apnoea, and Hypopnoea events.

The Obstructive and Central Apnoea models showed good performance. The models were also able to successfully identify Apnoea events even when there wasn’t an associated desaturation event, suggesting that the model had identified patterns or combinations of patterns in the other signals that were indicative of Apnoea.

CONCLUSIONS

The AI had some difficulty distinguishing Hypopnoea events from normal samples. This reduced the accuracy of the ‘All’ neural network despite it performing well on all of the Apnoea events, as Hypopnoea samples accounted for 66% of the total samples in the ‘All’ dataset. The reduced accuracy on Hypopnoeas was expected, as the physiological impact of a Hypopnoea is generally less severe than that of an Apnoea.

The samples were based on the labels given by sleep technicians during PSG studies at a sleep clinic. Due to time constraints the technicians are not always able to ensure perfect start and stop times when labelling events. As a result, the labels used in training the AI were sometimes inaccurate and improvements to the AI could also be achieved by further curating the dataset, as well as expanding it.

Using a trained artificial intelligence model, we have been able to correctly identify respiratory events using 20 second windows, with up to 85% accuracy. The AI provides a mechanism for automated respiratory condition detection from any device that monitors the aforementioned signals. In addition, machine learning allows for ongoing feedback and training, improving the detection accuracy as the dataset grows. The next step is to embed the respiratory AI into our smartphone application to run in conjunction with Alerte’s sleep wearable.

DECLARATION OF INTEREST: The authors of this poster hold a financial interest in Alerte Digital Health.