Interpreting Deep Neural Networks for Single-Lead ECG Arrhythmia Classification

Sricharan Vijayarangan, Balamurali Murugesan, Vignesh R, Preejith SP

Jayaraj Joseph and Mohansankar Sivaprakasam
Motivation

* Deep Learning methods for arrhythmia detection
  ○ Scales automated systems
  ○ Removes requirement for expert rules
  ○ Augments doctor’s ability

* Limitations
  ○ Black box
  ○ Unreliable

* Requirement
  ○ Correlation b/w model outputs and ECG input samples
  ○ Comparing visualizations with medical literature
Contribution

- Novel adaptation of CNN saliency visualization to 1D ECG signals
- Extension of the LSTM visualization procedure for ECG signals
- Rigorous analysis of the saliency maps
- Draw comparisons to traditional diagnosis as highlighted in medical literature
Problem formulation

Dataset $\rightarrow \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})\}$

Input ECG Signal $\rightarrow x^{(i)}$  
Labels $\rightarrow y^{(i)} \in \{0, 1, \ldots, 7\}$

CNN $\rightarrow z_1^{(i)} = F_1(x^{(i)}; \theta_1)$
LSTM $\rightarrow z_2^{(i)} = F_2(x^{(i)}; \theta_2)$

FC $\rightarrow z_3^{(i)} = F_3(z_1^{(i)} \| z_2^{(i)}; \theta_3)$

Softmax $\rightarrow p(z_3^{(i)})$
Rationale behind Architecture Choice

Three popular DL architectures in literature were compared for this specific 8 class classification problem.

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hannun et al. [2]</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>Zühlmann et al. [3]</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>Murugesan et al. [4]</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Murugesan et al.’s model is clearly the best for this classification task. Thus, it was chosen for the interpretability task.
Architecture Design

# Dataset Description

<table>
<thead>
<tr>
<th>Rhythm Types</th>
<th>MITDB</th>
<th>LTAFDB</th>
<th>LTDB</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>75013</td>
<td>10756</td>
<td>517402</td>
<td>603171</td>
</tr>
<tr>
<td>PVC</td>
<td>7121</td>
<td>1318</td>
<td>5137</td>
<td>13576</td>
</tr>
<tr>
<td>PAC</td>
<td>2542</td>
<td>14914</td>
<td>-</td>
<td>17456</td>
</tr>
<tr>
<td>AFIB</td>
<td>102</td>
<td>7241</td>
<td>-</td>
<td>7343</td>
</tr>
<tr>
<td>SVTA</td>
<td>22</td>
<td>3265</td>
<td>-</td>
<td>3287</td>
</tr>
<tr>
<td>SBR</td>
<td>-</td>
<td>11323</td>
<td>-</td>
<td>11323</td>
</tr>
<tr>
<td>LBBB</td>
<td>6580</td>
<td>-</td>
<td>-</td>
<td>6580</td>
</tr>
<tr>
<td>RBBB</td>
<td>5400</td>
<td>-</td>
<td>-</td>
<td>5400</td>
</tr>
</tbody>
</table>
Visualization of CNN

Activation of unit $k \rightarrow f_k(x)$  

Weight for class $C$ corresponding to Unit ‘$k$’ \[ w^c_k \]

I/p to Softmax \[ \sum_k w^c_k \sum_x f_k(x) \]

CAM \[ M_c(x) = \sum_k w^c_k f_k(x) \]

- CAM of vector length 48 is obtained.
- Upsampled to 720
LSTM Visualization

LSTM visualization \( (\psi) \)

Input ECG signal \( (x_{1:T}) \)

Weights of the saliency term \( \lambda_1 \)

Weights of the smoothing term \( \lambda_2 \)

Input to the LSTM network \( (\phi) \)

Mask \( (m_{1:T}) \)

\[
J = \arg \min_{m_{1:T}} \lambda_1 \| 1 - m_{1:T} \|_1 + \lambda_2 \sum_{t=1}^{T-1} |m_{t+1} - m_t| \\
+ s_c(\psi(\phi(x_{1:T}; m_{1:T})))
\]

\[
\phi(x_{1:T}; m_{1:T}) = m_{1:T} \odot x_{1:T} + k(1 - m_{1:T})
\]
Initially, $m_{1:T} = 0$

$\lambda_1, \lambda_2$ and learning rate are set to 1, 0.001 and 0.001 respectively.

Gradient update is done for 500 iterations.
ECG Visualization

<table>
<thead>
<tr>
<th></th>
<th>ECG Signal</th>
<th>CNN Visualization</th>
<th>LSTM Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>N:</td>
<td><img src="image1.png" alt="Signal" /></td>
<td><img src="image2.png" alt="CNN Visualization" /></td>
<td><img src="image3.png" alt="LSTM Visualization" /></td>
</tr>
<tr>
<td>PVC:</td>
<td><img src="image4.png" alt="Signal" /></td>
<td><img src="image5.png" alt="CNN Visualization" /></td>
<td><img src="image6.png" alt="LSTM Visualization" /></td>
</tr>
<tr>
<td>PAC:</td>
<td><img src="image7.png" alt="Signal" /></td>
<td><img src="image8.png" alt="CNN Visualization" /></td>
<td><img src="image9.png" alt="LSTM Visualization" /></td>
</tr>
<tr>
<td>AF:</td>
<td><img src="image10.png" alt="Signal" /></td>
<td><img src="image11.png" alt="CNN Visualization" /></td>
<td><img src="image12.png" alt="LSTM Visualization" /></td>
</tr>
<tr>
<td>SVTA:</td>
<td><img src="image13.png" alt="Signal" /></td>
<td><img src="image14.png" alt="CNN Visualization" /></td>
<td><img src="image15.png" alt="LSTM Visualization" /></td>
</tr>
<tr>
<td>SBR:</td>
<td><img src="image16.png" alt="Signal" /></td>
<td><img src="image17.png" alt="CNN Visualization" /></td>
<td><img src="image18.png" alt="LSTM Visualization" /></td>
</tr>
<tr>
<td>LBBB:</td>
<td><img src="image19.png" alt="Signal" /></td>
<td><img src="image20.png" alt="CNN Visualization" /></td>
<td><img src="image21.png" alt="LSTM Visualization" /></td>
</tr>
<tr>
<td>RBBB:</td>
<td><img src="image22.png" alt="Signal" /></td>
<td><img src="image23.png" alt="CNN Visualization" /></td>
<td><img src="image24.png" alt="LSTM Visualization" /></td>
</tr>
</tbody>
</table>
Confusion Matrix of Predictions

Interpretability w.r.t epochs no
Conclusion and Future scope

- A novel adaptation of visualization techniques of CNN and LSTM for ECG signals was proposed.
- Visualizations were observed to line up with the clinical literature in ECG interpretation.
- Extension to other arrhythmia classes.
- Extension to entire arrhythmia records.
- Exploring Explainability.
Thank you

Contact - sricharanv@htic.iitm.ac.in