Zhao, Zheng; Särkkä, Simo; Bahrami Rad, Ali

Spectro-Temporal ECG Analysis for Atrial Fibrillation Detection

Published in:
2018 IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2018

DOI:
10.1109/MLSP.2018.8517085

Published: 01/01/2018

Document Version
Peer reviewed version

Please cite the original version:
This work was supported by Business Finland.
We now assume that the coefficients depend on time, and we put a Gaussian process priors on them:

\[ a_j(t) \sim \mathcal{GP}(0, k_j^a(t, t')), \]
\[ b_j(t) \sim \mathcal{GP}(0, k_j^b(t, t')). \]  
(2)

As shown in [23, 24], provided that the covariance functions are stationary, we can express the Gaussian processes as solutions to linear stochastic differential equations (SDEs). In this paper we choose covariance functions to have the form

\[ k_j^a(t, t') = (s_j^a)^2 \exp(-\lambda_j^a |t - t'|), \]
\[ k_j^b(t, t') = (s_j^b)^2 \exp(-\lambda_j^b |t - t'|), \]  
(3)

where \( s_j^a, s_j^b > 0 \) are scale parameters and \( \lambda_j^a, \lambda_j^b > 0 \) are the inverses of the time constants (length scales) of the processes. The state-space representations (which are scalar in this case) are then given as

\[ da_j = -\lambda_j^a a_j dt + dW_j^a, \]
\[ db_j = -\lambda_j^b b_j dt + dW_j^b, \]  
(4)

where \( W_j^a, W_j^b \) are Brownian motions with suitable diffusion coefficients \( q_j^a, q_j^b \). We can also solve the equations at discrete time steps (see, e.g., [25]) as

\[ a_j(t_k) = \psi_j^a b_j(t_{k-1}) + b_j^a, \]
\[ b_j(t_k) = \psi_j^b b_j(t_{k-1}) + b_j^b, \]  
(5)

where \( \psi_j^a = \exp(-\lambda_j^a (t_k - t_{k-1})), \psi_j^b = \exp(-\lambda_j^b (t_k - t_{k-1})), w_j^a \sim \mathcal{N}(0, \Sigma_{j,k}^a), w_j^b \sim \mathcal{N}(0, \Sigma_{j,k}^b), \Sigma_{j,k}^a = q_j^a (1 - \exp(-2\lambda_j^a (t_k - t_{k-1}))), \Sigma_{j,k}^b = q_j^b (1 - \exp(-2\lambda_j^b (t_k - t_{k-1}))). \)

Let us now assume that we obtain noisy measurements of the Fourier series (1) and times \( t_1, t_2, \ldots \). What we can now do is to stack all the coefficients into the state \( x = [a_0, a_1, \ldots, a_M, b_1, b_2, \ldots, b_M, k_1, \ldots, k_M] \). \( \mathbf{H}_k = [1, \sin(2\pi f_0 t_1), \sin(2\pi f_0 t_2), \ldots, \sin(2\pi f_M t_1), \cos(2\pi f_0 t_1), \ldots, \cos(2\pi f_M t_1)] \), which gives

\[ \mathbf{H}_k x = a_0 + \sum_{j=1}^M [a_j \cos(2\pi j f_0 t) + b_j \sin(2\pi j f_0 t)] = z(t_k). \]  
(6)

The discrete-time dynamic model (5) can be written as

\[ \mathbf{x}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{q}_k \]  
(7)

where \( \mathbf{A}_k \) contains the terms \( \psi_j^a b_j \) and \( \psi_j^b b_j \) on the diagonal and \( \mathbf{q}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \) where \( \mathbf{Q}_k \) contains the terms \( \Sigma_{j,k}^a \) and \( \Sigma_{j,k}^b \) on the diagonal.

If we assume that we actually measure (6) with additive Gaussian measurement noise \( r_k \sim \mathcal{N}(0, R) \), then we can express the measurement model as

\[ y_k = \mathbf{H}_k \mathbf{x}_k + r_k. \]  
(8)

The Kalman filter for this problem then consists of the following forward recursion (for \( k = 1, \ldots, N \)):

\[ m_k^- = \mathbf{A}_k \mathbf{m}_{k-1}, \]
\[ P_k^- = \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^\top + \mathbf{Q}_k, \]
\[ S_k = \mathbf{H}_k \mathbf{P}_{k}^- \mathbf{H}_k^\top + \mathbf{R}, \]
\[ \mathbf{K}_k = \mathbf{P}_{k}^- \mathbf{H}_k^\top / S_k, \]
\[ \mathbf{m}_k = \mathbf{m}_k^- + \mathbf{K}_k \left( y_k - \mathbf{H}_k \mathbf{m}_k^- \right), \]
\[ \mathbf{P}_k = \mathbf{P}_{k}^- - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top, \]  
(9)

and the RTS smoother the following backward recursion (for \( k = N - 1, \ldots, 1 \)):

\[ \mathbf{G}_k = \mathbf{P}_k \mathbf{A}_{k+1}^\top \left[ \mathbf{P}_{k+1}^- \right]^{-1}, \]
\[ \mathbf{m}_k^- = \mathbf{m}_k + \mathbf{G}_k \left( \mathbf{m}_{k+1}^- - \mathbf{m}_{k+1} \right), \]
\[ \mathbf{P}_k^- = \mathbf{P}_k + \mathbf{G}_k \left[ \mathbf{P}_{k+1}^- - \mathbf{P}_{k+1} \right] \mathbf{G}_k^\top. \]  
(10)

The final posterior distributions are then given as:

\[ p(x_k | y_1:N) = \mathcal{N}(x_k | \mathbf{m}_k^\ast, \mathbf{P}_k^\ast). \]  
(11)

The magnitude of the sinusoidal with frequency \( f_j = j f_0 \) at time step \( k \) can then be computed by extracting the elements corresponding to \( a_j(t_k) \) and \( b_j(t_k) \) from the mean vector \( \mathbf{m}_k^\ast \) as

\[ |S|_{j,k} = \sqrt{a_j^2(t_k) + b_j^2(t_k)}. \]  
(12)

From now, matrix \( \mathbf{S} \) is called spectro-temporal data matrix.

3. ATRIAL FIBRILLATION DETECTION USING SPECTRO-TEMPORAL ECG CLASSIFICATION

3.1. Processing chain

The processing chain of the proposed scheme is illustrated in Fig. 1. The raw ECG is first segmented, and then the spectro-temporal data matrix of each segment is computed using (12). The resulting spectro-temporal data matrices are then averaged and normalized to generate fixed-length spectro-temporal feature matrix. Finally, the 2D feature matrix (spectro-temporal image) is fed into a deep convolutional neural network (CNN) for classification.

In this work, we use a modified version of Pan-Tompkins algorithm for QRS detection. The original Pan-Tompkins algorithm [26]

![Fig. 1. Processing chain of ECG classification](image-url)
is sensitive to burst noise, and it easily misinterprets noise with R peak. To address this limitation at least partially, we slightly modify the original algorithm such that it iteratively checks the number of detected R peaks and if that number is smaller than a threshold, it ignores the detected R peaks and their neighbourhood samples in the ECG signal, and again applies the Pan-Tompkins algorithm on the rest of the signal. In this way, if there are few instances with high-amplitude burst noise, our algorithms can handle those.

As the next step we have a representation averaging procedure that aims to produce an input for deep CNNs classifier by averaging the fixed length spectral blocks containing three QRS complexes. If \( z = [z_1, z_2, \ldots, z_N]^\top \in \mathbb{R}^N \) is the original ECG signal and \( \bar{p}_i \in \{1, 2, \ldots, N\} \) is the position of \( i \)th R peak in \( z \), then \( \bar{p} = [\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_D] \) holds the positions of all R peaks in \( z \). Now, we associate each \( \bar{p}_i \), \( i \in \{2, \ldots, D - 1\} \), to a segment of \( z \) such that it potentially covers three adjacent QRS complexes. To do so, we collect \( \beta \) samples before and after each \( \bar{p}_i \). Following this procedure, the ECG segment associated to \( i \)th R peak can be extracted from \( z \) as \( z^{(i)} = [z_{\bar{p}_i - \beta}, \ldots, z_{\bar{p}_i + \beta}] \), and using equation (12), the spectro-temporal data matrix corresponding to this ECG segment is \( S^{(i)} \in \mathbb{R}^{M \times (2\beta + 1)} \) where \( M \) and \( 2\beta + 1 \) are frequency and time steps, respectively.

The spectro-temporal feature matrix \( S^\top \) is obtained by averaging over all spectro-temporal data matrices and multiplying with their maximum mask:

\[
S^\top = \frac{1}{D - 2} \max_{2 \leq t \leq D - 1} S^{(i)}.
\]

The choice of parameter \( \beta \) is important, as it regulates the length of output and how much takes into average. Usually, \( \beta \) should at least covers three QRS complexes or more for good evidence of R-R interval. The reason for adding max operation in averaging is that it could help preserving intricate details of spectro-temporal data. Examples on representation averaging for four classes of ECG signals are shown in Fig. 2.

### 3.2. Time–Frequency Analysis

Although for ECG classification, we employ the spectro-temporal representation described in Section 2, other standard time–frequency analysis methods are also examined for the sake of comparison. We use magnitude of CWT, magnitude of STFT, and square root of non-logarithmic power spectral density using Burg autoregressive model (BurgAR) \([27]\) of ECG signal.

Fig. 3 shows the spectro-temporal representation of an ECG segment by different methods. (a) is the original signal of Rec.3246 in CinC 2017. We control frequency range \((M)\) and smoothing option of Kalman method, as shown in subfigures (b), (c), and (d). Subfigure (e) shows result by the original method in \([21]\). Subfigures (f), (g), and (h) show STFT, CWT, and BurgAR methods results respectively, where we apply 11 length 10 overlapping Hamming windows on STFT, BurgAR, and CWT (with Morse wavelet). For our proposal, we choose 10 for length scale \( \lambda \) and 1 for variance of both process and measurement noise \( R \) and \( q \).

If we compare subfigures (c), (f), (g), and (h), we can find three advantages of Kalman method over STFT, BurgAR and CWT: the result is more smooth and it has higher and more unified resolution on both time and frequency. For STFT and BurgAR, the resolution is confined by window length selection. CWT solves this by replacing window with wavelet, but due to uncertainty principle of signal processing, the required resolution in time and frequency cannot be met simultaneously. We can see in (g) that the time resolution is very low in low frequency bands. Our approach models the time-varying Fourier series coefficients of signal in state-space, which achieves observation-wise spectrogram estimation.

### 3.3. Densely Connected Convolutional Networks

In recent years, deep learning techniques especially various convolutional neural networks, emerge as dominant methods for image classification. However, one flaw is that the information during training, principally the gradient, may disappear if the network is exceedingly deep (with many layers), which is usually called “vanishing gradient” \([28]\). Generally, this root problem can be alleviated by several basic ways, for instance, with layer-wise and pre-training, or with a properly selected activation function. Densely connected convolutional networks (DenseNet) \([16]\), who won the 2017 best paper award of CVPR, provide state-of-the-art performance without
DenseNet can be seen as a refined version of deep residual networks (ResNet) [29], where the former one introduces explicit connection degradation or over-fitting even when stacked by hundred of layers. The DenseNet we implement here, which we refer as Dense18\(^+\), is slightly different from the original proposal, where we employ both max and average global pooling on last layer as shown in Fig. 5. Each dense block contains four \(3 \times 3\) convolutional layers, with growth rate of 48.

### 4. EXPERIMENTS

#### 4.1. ECG Dataset and Evaluation Metrics

We have conducted experiments on the PhysioNet/CinC 2017 dataset [17] to evaluate the performance of the proposed method. The dataset contains 8528 short ECG recordings (9s to 60s) at 300Hz sampling rate. For model assessment we use stratified 10-fold cross-validation. The detailed performance is evaluated using a 4-class confusion matrix, where the diagonal entries are the correct classifications and the off-diagonal entries are the incorrect classifications. This confusion matrix is the result of stacking 10 confusion matrices of the test data in the 10-fold cross-validation. In addition, the F1 score,

\[
F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}},
\]

for each class is calculated to summarize the performance of the proposed method for that specific class: Normal \(F1_N\), AF \(F1_A\), Others \(F1_O\), and Noisy \(F1_\text{Noisy}\). Finally, the overall performance of the proposed algorithm is evaluated using the suggested evaluation metric by PhysioNet/CinC 2017 [17]:

\[
F1_{\text{overall}} = \frac{1}{3}(F1_N + F1_A + F1_O).
\]

#### 4.2. Results

We first compare the results of our proposal (Kalman) and other spectro-temporal representation methods (CWT, STFT, and BurgAR) upon a same classifier Dense18\(^+\). The settings for spectrogram estimation we choose here are the same as described in Section 3. All spectro-temporal feature matrices (images) are then unifiedly resized (down-sample by local averaging) to 50 \(\times\) 50 for Dense18\(^+\).

As shown in rows (1)–(4) of Table 1, the proposed Bayesian spectro-temporal method achieves an overall F1 score of 80.17, which surpasses STFT (77.79), CWT (79.55), and BurgAR (77.95) for ECG classification. It also has the highest F1 scores for detection of Normal, AF, and other rhythms: 88.80, 79.64, and 72.08, respectively. In addition, the proposed method has the lowest cross-validation standard deviation \(\text{Std}_{F1}\) 1.06, suggesting higher robustness and reliability.

The detailed performance of all four methods (i.e., Kalman, CWT, STFT, and BurgAR) are reported in four confusion matrices in Fig. 6. Each confusion matrix is row-wise normalized. The diagonal entries show the Recall of each rhythm and off-diagonal entries show the misclassification rates. For example, the first row of the first confusion matrix shows 90.6% of normal rhythms are correctly classified as normal, but 0.4%, 8.0%, and 0.9% are incorrectly classified as AF, Others, and Noisy.

#### 4.3. Discussion

Let us now discuss the reasons why the Kalman filter based approach produces better results in the classification. One way to study the resulting classifier is to investigate its first convolutional layer which corresponds to the (dominant) features that the deep CNN has learned [31]. The layer is shown in Fig. 7. The figure shows that the network has larger activation on shape, edge and intensity of “peaks” and more importantly, the details of background. The “peaks” and details are very crucial for AF detection, because they
can respectively represent R-R interval and shape of P wave. In comparison to CWT, STFT, and BurgAR, the background details are better preserved in the Kalman method. Furthermore, the Kalman approach works well with non-stationary signals that we have in the AF rhythm.

In Fig. 8, we show how features are correlated by performing Variational Autoencoder (VAE) [32] and t-Stochastic Neighbour Embedding (t-SNE) [33] visualization on the last concatenate layer before Softmax classifier of Dense18+ . We can find that after the training of deep CNNs, the learnt features are well embedded and correlated by classes in high dimensional feature space (mapped into two dimension). Although AF and Normal rhythm classes are well separated, the Other and Noisy classes still have strong overlap with them, which can also be seen in the confusion matrices in Fig. 6. We assume that the representation averaging procedure may well represent AF and Normal rhythms, however, it faints the differences from Other and Noisy classes, which causes low performance in Other and Noisy classes. The classes of the Kalman approach seem have less overlap compared to CWT, STFT, and BurgAR.

In Table 1, rows (4)–(6) compare the performance when applying Kalman method with other two different deep CNNs architectures: Dense18 and Res18, which both have the equivalent depth (convolutional layers) with Dense18+ in this paper. The results state that DenseNet has a better overall performance than ResNet in AF detection, and our modification on last pooling layer (Dense18+) improves the performance ($F_{overall}$) by 1.64 percentage points to original Dense18 networks.

We also compared our performance with [15, 30], where the authors adopted a similar approach for AF detection, that is, spectrogram and deep CNNs, during 2017 PhysioNet/CinC Challenge. The results show that our combination using spectro-temporal analysis and DenseNet outperforms them. Although our method is in line with the state-of-the-art algorithms, the winners of the challenge used fine-tuned hand-crafted features which also reflect the expert knowledge to achieve the cutting-edge performance (83%). In the future we will investigate hybrid methods which incorporate expert

**Table 1.** 10-fold cross-validation results using different spectrogram estimation methods and deep CNNs architectures.

<table>
<thead>
<tr>
<th>Methods</th>
<th>$F_{IN}$</th>
<th>$F_{IA}$</th>
<th>$F_{IO}$</th>
<th>$F_{IA}$/</th>
<th>$F_{overall}$</th>
<th>Std $F_{IA}$</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) CWT + Dense18</td>
<td>88.06</td>
<td>75.23</td>
<td>70.03</td>
<td>52.67</td>
<td>77.79</td>
<td>1.41</td>
<td>79.38</td>
</tr>
<tr>
<td>(2) Noisy</td>
<td>88.77</td>
<td>78.08</td>
<td>71.79</td>
<td>53.25</td>
<td>79.55</td>
<td>1.39</td>
<td>80.73</td>
</tr>
<tr>
<td>(3) Noisy</td>
<td>88.11</td>
<td>76.24</td>
<td>69.49</td>
<td>55.91</td>
<td>77.95</td>
<td>1.51</td>
<td>79.23</td>
</tr>
<tr>
<td>(4) Noisy</td>
<td>88.80</td>
<td>79.64</td>
<td>72.08</td>
<td>51.78</td>
<td>80.17</td>
<td>1.06</td>
<td>81.33</td>
</tr>
<tr>
<td>(5) Kalman + Dense18</td>
<td>88.16</td>
<td>76.61</td>
<td>70.81</td>
<td>49.21</td>
<td>78.53</td>
<td>1.14</td>
<td>80.02</td>
</tr>
<tr>
<td>(6) Kalman + Res18[29]</td>
<td>87.19</td>
<td>74.98</td>
<td>68.31</td>
<td>47.01</td>
<td>76.83</td>
<td>1.08</td>
<td>78.04</td>
</tr>
<tr>
<td>(7) [15]</td>
<td>87.8</td>
<td>79.0</td>
<td>70.1</td>
<td>65.3</td>
<td>79.0</td>
<td>N/A</td>
<td>81.2</td>
</tr>
<tr>
<td>(8) [30]</td>
<td>87</td>
<td>80</td>
<td>68</td>
<td>N/A</td>
<td>78</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Fig. 5.** Structure of Dense18+ in this paper.

**Fig. 6.** Normalized confusion matrix on different methods.

**Fig. 7.** Feature-map (Left 16 columns) and activation (right 16 columns) visualization of first convolutional layer on Rec. 1005 (AF). From top to bottom, every 4 rows are Kalman, CWT, BurgAR and STFT respectively.
knowledge to the deep learning models.

5. CONCLUSION

In this paper, we present a new spectro-temporal analysis method by assuming the time-varying Fourier coefficients of signal have Gaussian process priors. We express the solution in linear state-space and use a Bayesian Kalman filter/smooother for parameter estimation. Combining the aforementioned spectro-temporal representation with CNNs for ECG classification outperforms other time-frequency analysis methods (i.e., STFT, CWT, and BurgAR) with the same classifier for AF detection. The proposed method provides the classification performance of 80.2% for overall F1 score.

6. REFERENCES