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

Spectro-Temporal ECG Analysis for Atrial Fibrillation Detection

SPECTRO-TEMPORAL ECG ANALYSIS FOR ATRIAL FIBRILLATION DETECTION

Zheng Zhao, Simo Särkkä, and Ali Bahrami Rad

Aalto University, Finland
Department of Electrical Engineering and Automation

ABSTRACT

This article is concerned with spectro-temporal (i.e., time varying spectrum) analysis of ECG signals for application in atrial fibrillation (AF) detection. We propose a Bayesian spectro-temporal representation of ECG signal using state-space model and Kalman filter. The 2D spectro-temporal data are then classified by a densely connected convolutional networks (DenseNet) into four different classes: AF, non-AF normal rhythms (Normal), non-AF abnormal rhythms (Others), and noisy segments (Noisy). The performance of the proposed algorithm is evaluated and scored with the PhysioNet/Computing in Cardiology (CinC) 2017 dataset. The experiment results shows that the proposed method achieves the overall F1 score of 80.2%, which is in line with the state-of-the-art algorithms. In addition, the proposed spectro-temporal estimation approach outperforms standard time-frequency analysis methods, that is, short-time Fourier transform, continuous wavelet transform, and autoregressive spectral estimation for AF detection.

Index Terms— Atrial fibrillation, deep learning, Kalman filter, state-space model, spectrogram estimation

1. INTRODUCTION

Atrial fibrillation (AF) is a type of cardiac rhythm disturbance (arrhythmia) which can lead to blood clots, stroke, heart failure, or death. AF is the most common cardiac arrhythmia affecting around 33.5 million individuals worldwide in 2010 [1]. It is also estimated that the number of patients with AF in Europe Union will be 14–17 million by 2030 [2]. AF is defined as chaotic electrical activity of atrial muscle fibers. During AF, atrioventricular (AV) node may receive more than 500 impulses per minute, from which only occasional impulses can pass through at variable rate, resulting irregular ventricular response [3]. Manifestations of AF on ECG are the absence of P-wave and irregular RR intervals [4].

Aiming to detect AF automatically, various algorithms have been developed [5–9]. In addition to traditional approaches, recent deep learning (DL) techniques also provide a promising end-to-end classification for ECG signals. Unlike traditional approaches, one of the most significant advantage of using deep learning for classification is that hand-crafted features are no longer needed, because deep neural networks have the ability of learning the inherent features when provided with a sufficient training data [10]. Whilst surprisingly, the combination of AF and deep learning has just begun in past two years (see, e.g., [11–14]).

However, within the most of the previous studies, only few have resorted to use spectrogram for AF detection. It is hard to select handmade features from 2D data for traditional methods, and thus in

this case DL models are advantageous. Several studies [13, 15] have endeavoured DL for AF detection in spectral domain, but the use of spectral estimation methods such as short-time Fourier transform (STFT) or continuous wavelet transform (CWT), may drop momentous information during the transformation, and produce less informative images. Furthermore, previous studies (e.g. [12–14]) are performed on almost clean dataset selected from small-scale number of patients and focusing on binary classification of ECG into AF and non-AF rhythms, which is usually not practical in production environment.

The contributions of the paper are: 1) We propose an extended spectrogram estimation method by modeling signal in state-space and use a Kalman filter and smoother for Bayesian spectral estimation; 2) we leverage state-of-the-art densely connected convolutional networks [16] for AF detection using the proposed presentation; 3) we evaluate the method using PhysioNet/CinC 2017 dataset [17], which is considered to be a challenging dataset which resembles practical applications.

The paper is structured as follows: In Section 2, we introduce the spectro-temporal method for ECG signals analysis. In Section 3, we use the proposed estimation method on AF detection together with deep convolutional networks. In Section 4, we compare and discuss experiments results, followed by conclusion in Section 5.

2. SPECTRO-TEMPORAL ECG ANALYSIS

The spectro-temporal signal analysis is an effective and powerful approach in many fields [18–20]. In this section we present a Bayesian spectro-temporal ECG analysis approach, which is an extension of the Bayesian spectrum estimation method of Qi et al. [21].

We model the time varying spectrum with a state-space model and use Bayesian procedure (i.e., Kalman filter and smoother) for its estimation [21, 22]. One of the significant advantage is that it can be applied on both evenly and unevenly sampled signals [21] and does not need stationarity guarantee nor windowing. However, as we show here, it can also be combined with state-space methods for Gaussian processes [23, 24].

Recall that any periodic signal with fundamental frequency f_0 can be expanded into a Fourier series

$$z(t) = a_0 + \sum_{j=1}^{M} \left[a_j \cos(2\pi j f_0 t) + b_j \sin(2\pi j f_0 t) \right], \quad (1)$$

where the exact series if obtained with $M \to \infty$, but for sampled signals it is sufficient to consider finite series. This stationary model is indeed the underlying model in the STFT approach. What STFT effectively does, is that it does least squares fit of the coefficients $\{a_j, b_j : j = 1, \ldots, M\}$ at each window separately.

We now assume that the coefficients depend on time, and we put a Gaussian process priors on them:

$$
\begin{aligned}
a_j(t) &\sim \mathcal{GP}(0, k_j^a(t, t')), \\
b_j(t) &\sim \mathcal{GP}(0, k_j^b(t, t')).
\end{aligned}
\tag{2}
$$

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

$$
\begin{aligned}
k_j^a(t, t') &= (s_j^a)^2 \exp(-\lambda_j^a |t - t'|), \\
k_j^a(t, t') &= (s_j^b)^2 \exp(-\lambda_j^b |t - t'|),
\end{aligned}
\tag{3}
$$

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

$$
\begin{aligned}
da_j &= -\lambda_j^a \, a_j \, dt + dW_j^a, \\
db_j &= -\lambda_j^b \, b_j \, dt + dW_j^b,
\end{aligned}
\tag{4}
$$

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

$$
\begin{aligned}
a_j(t_k) &= \psi_{jk}^a \, a_j(t_{k-1}) + w_{jk}^a, \\
b_j(t_k) &= \psi_{jk}^b \, b_j(t_{k-1}) + w_{jk}^b,
\end{aligned}
\tag{5}
$$

where $\psi_{jk}^a = \exp(-\lambda_j^a (t_k - t_{k-1})), \psi_{jk}^b = \exp(-\lambda_j^b (t_k - t_{k-1})), w_{jk}^a \sim \mathcal{N}(0, \Sigma_{jk}^a), w_{jk}^a \sim \mathcal{N}(0, \Sigma_{jk}^b), \Sigma_{jk}^a = q_j^a (1 - \exp(-2\lambda_j^a (t_k - t_{k-1}))),$ and $\Sigma_{jk}^b = q_j^b (1 - \exp(-2\lambda_j^b (t_k - t_{k-1})))$.

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 $\mathbf{x} = [a_0, a_1, ..., a_M, b_1, b_2, ..., b_M]^T$, $\mathbf{H}_k = [1, \sin(2\pi f_0 t_k), \ldots, \sin(2\pi M f_0 t_k), \cos(2\pi f_0 t_k), \ldots, \cos(2\pi f_M t_j)]$, which gives

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

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

$$
\mathbf{x}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{q}_k
\tag{7}
$$

where $\mathbf{A}_k$ contains the terms ψ_{jk}^a and ϕ_{jk}^b on the diagonal and $\mathbf{q}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ where $\mathbf{Q}_k$ contains the terms Σ_{jk}^a and Σ_{jk}^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

$$
y_k = \mathbf{H}_k \mathbf{x}_k + r_k.
\tag{8}
$$

Equations (7) and (8) define a linear state-space model where we can perform exact Bayesian estimation using Kalman filter and Rauch–Tung–Striebel (RTS) smoother [22]. In the original paper [21], the state vectors $\mathbf{x}_1, ..., \mathbf{x}_N$ are assumed to perform random walk, but here the key insight is to use a more general Gaussian process which introduces a finite time constant to the problem.

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

$$
\begin{aligned}
\mathbf{m}_k^- &= \mathbf{A}_k \, \mathbf{m}_{k-1}, \\
\mathbf{P}_k^- &= \mathbf{A}_k \, \mathbf{P}_{k-1} \, \mathbf{A}_k^\top + \mathbf{Q}_k, \\
S_k &= \mathbf{H}_k \, \mathbf{P}_k^- \, \mathbf{H}_k^\top + R, \\
\mathbf{K}_k &= \mathbf{P}_k^- \, \mathbf{H}_k^\top / S_k, \\
\mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k \left(y_k - \mathbf{H}_k \, \mathbf{m}_k^-\right), \\
\mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k \, S_k \, \mathbf{K}_k^\top,
\end{aligned}
\tag{9}
$$

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

$$
\begin{aligned}
\mathbf{G}_k &= \mathbf{P}_k \, \mathbf{A}_{k+1}^\top \, [\mathbf{P}_{k+1}^-]^{-1}, \\
\mathbf{m}_k^s &= \mathbf{m}_k + \mathbf{G}_k \, [\mathbf{m}_{k+1}^s - \mathbf{m}_{k+1}^-], \\
\mathbf{P}_k^s &= \mathbf{P}_k + \mathbf{G}_k \, [\mathbf{P}_{k+1}^s - \mathbf{P}_{k+1}^-] \, \mathbf{G}_k^\top.
\end{aligned}
\tag{10}
$$

The final posterior distributions are then given as:

$$
p(\mathbf{x}_k \mid y_{1:N}) = \mathcal{N}(\mathbf{x}_k \mid \mathbf{m}_k^s, \mathbf{P}_k^s). \quad k = 1, \ldots, N.
\tag{11}
$$

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 $\hat{a}_j(t_k)$ and $\hat{b}_j(t_k)$ from the mean vector $\mathbf{m}_k^s$:

$$
[\mathbf{S}]_{j,k} = \sqrt{\hat{a}_j^2(t_k) + \hat{b}_j^2(t_k)}.
\tag{12}
$$

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]
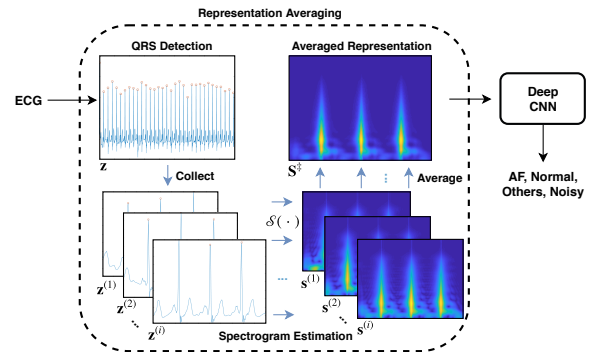

Fig. 1. Processing chain of ECG classification

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 $\mathbf{z} = [z_1, z_2, ..., z_N]^\top \in \mathbb{R}^N$ is the original ECG signal and $\bar{p}_i \in \{1, 2, \cdots, N\}$ is the position of ith R peak in $\mathbf{z}$, then $\bar{\mathbf{p}} = [\bar{p}_1, \bar{p}_2, ..., \bar{p}_D]^\top$ holds the positions of all R peaks in $\mathbf{z}$. Now, we associate each $\bar{p}_i$, $i \in \{2, \cdots, D-1\}$, to a segment of $\mathbf{z}$ such that it potentially covers three adjacent QRS complexes. To do so,

(a) Normal. Rec. 1012

(b) Normal. Rec. 1012

(c) Atrial Fibrillation. Rec. 3246

(d) Atrial Fibrillation. Rec. 3246

(e) Others. Rec. 1037

(f) Others. Rec. 1037

(g) Noise. Rec. 1006

(h) Noise. Rec. 1006

Fig. 2. Results of representation averaging (right sides) on four types of ECG signals (left sides), using proposed spectral analysis method. Red circles indicate detected R peaks.

we collect β samples before and after each $\bar{p}_i$. Following this procedure, the ECG segment associated to ith R peak can be extracted from $\mathbf{z}$ as $\mathbf{z}^{(i)} = [z_{\bar{p}_i - \beta}, \cdots, z_{\bar{p}_i}, \cdots, z_{\bar{p}_i + \beta}]^\top$, and using equation (12), the spectro-temporal data matrix corresponding to this ECG segment is $\mathbf{S}^{(i)} \in \mathbb{R}^{M \times (2\beta+1)}$ where M and $2\beta + 1$ are frequency and time steps, respectively.

The spectro-temporal feature matrix $\mathbf{S}^\ddagger$ is obtained by averaging over all spectro-temporal data matrices and multiplying with their maximum mask:

$$\mathbf{S}^\ddagger = \frac{\sum_{i=2}^{D-1} \mathbf{S}^{(i)}}{D-2} \circ \max_{2 \leq i \leq D-1} \mathbf{S}^{(i)}. \tag{13}$$

The choice of parameter β is important, as it regulates the length of output and how much takes into average. Usually, β 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.
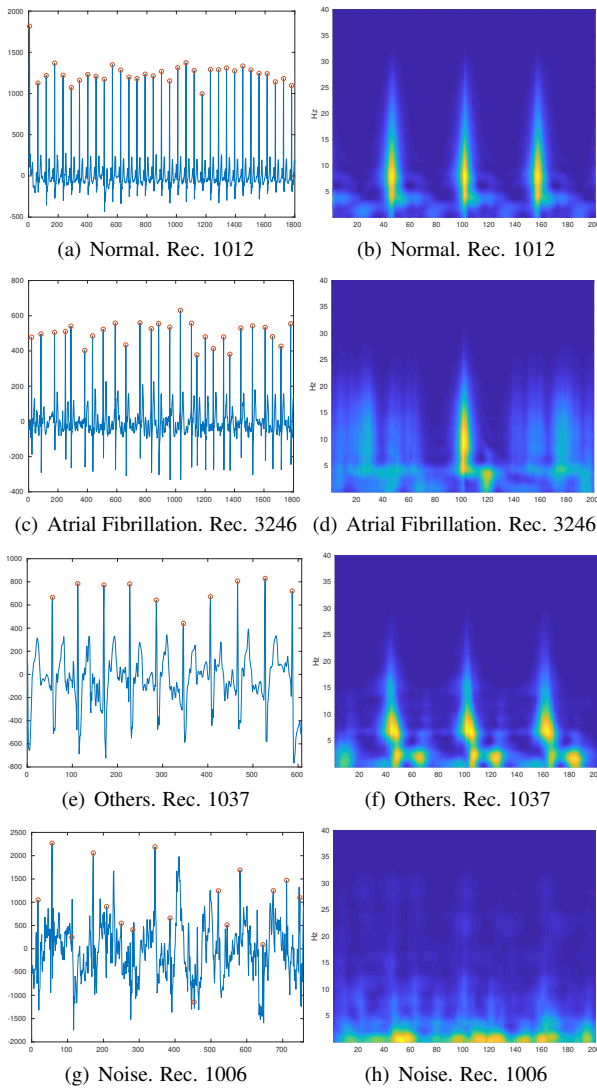
3.2. Time–Frequency Analysis

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 Hanning windows on STFT, BurgAR, and CWT (with Morse wavelet). For our proposal, we choose 10 for length scale λ, and 1 for variance of both process and measurement noise R and q.
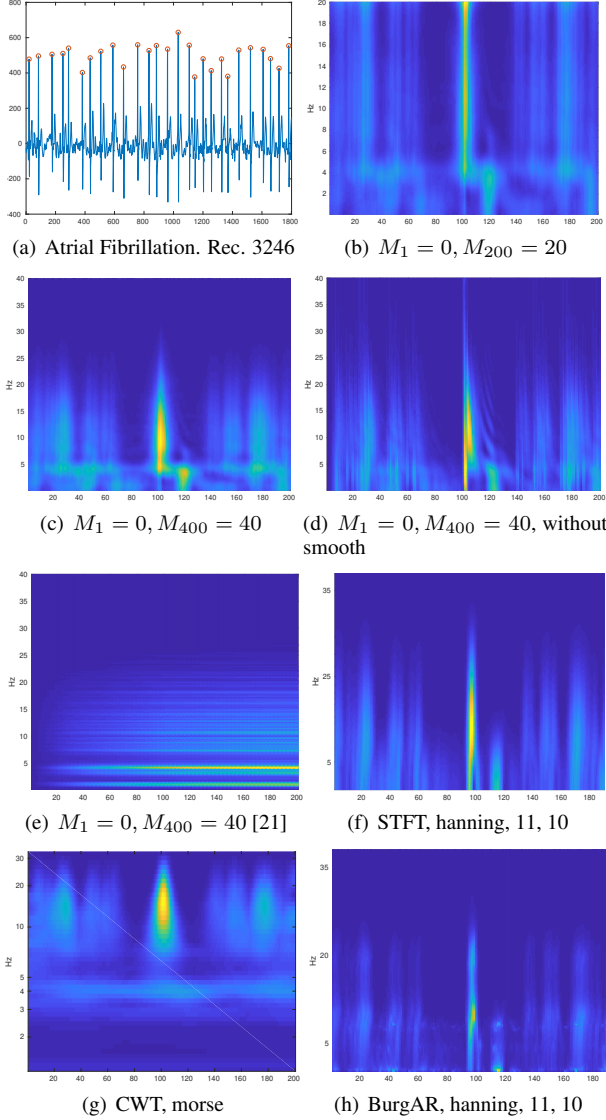
If we compare subfigures (c), (f), (g), and (h), we can find three advantages of Kalman method over STFT, BurgAR and CWT: the result is more smooth and it has higher and more unified resolution on both time and frequency. For STFT and BurgAR, the resolution is confined by window length selection. CWT solves this by replacing window with wavelet, but due to uncertainty principle of signal processing, the required resolution in time and frequency can not 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

(a) Atrial Fibrillation. Rec. 3246

(b) $M_1 = 0, M_{200} = 20$

(c) $M_1 = 0, M_{400} = 40$

(d) $M_1 = 0, M_{400} = 40$, without smooth

(e) $M_1 = 0, M_{400} = 40$ [21]

(f) STFT, hanning, 11, 10

(g) CWT, morse

(h) BurgAR, hanning, 11, 10

**Fig. 3**. Comparison of different spectrogram estimation methods on Rec. 3246.

degradation or over-fitting even when stacked by hundred of layers. DenseNet can be seen as a refined version of deep residual networks (ResNet) [29], where the former one introduces explicit connection on every two layers in a dense block, as shown in Fig. 4.

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 3×3 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
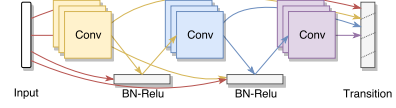
Fig. 4. Dense Block: each of of the convolutional layers takes all of the preceding outputs as input.

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,

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (14)$$

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 ($F1_\sim$). Finally, the overall performance of the proposed algorithm is evaluated using the suggested evaluation metric by PhysioNet/CinC 2017 [17]:

$$F1_{\text{overall}} = \frac{1}{3}(F1_N + F1_A + F1_O). \quad (15)$$

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 50×50 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 (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
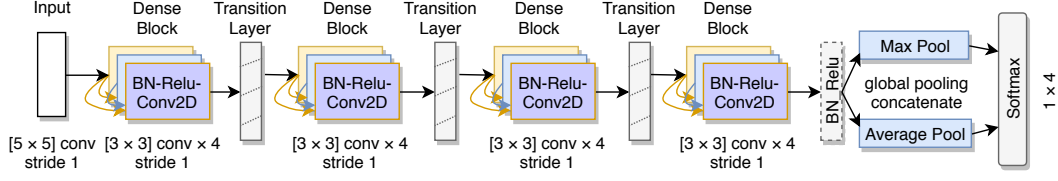
Fig. 5. Structure of Dense18$^+$ in this paper.

	Methods	$F1_N$	$F1_A$	$F1_O$	$F1_\sim$	$F1_{overall}$	Std_{F1}	Precision
(1)	STFT + Dense18$^+$	88.06	75.23	70.03	52.67	77.79	1.41	79.38
(2)	CWT + Dense18$^+$	88.77	78.08	71.79	53.25	79.55	1.39	80.73
(3)	BurgAR + Dense18$^+$	88.11	76.24	69.49	**55.91**	77.95	1.51	79.23
(4)	Kalman + Dense18$^+$	**88.80**	**79.64**	**72.08**	51.78	**80.17**	**1.06**	**81.33**
(5)	Kalman + Dense18	88.16	76.61	70.81	49.21	78.53	1.14	80.02
(6)	Kalman + Res18[29]	87.19	74.98	68.31	47.01	76.83	1.08	78.04
(7)	Martin[15]	87.8	79.0	70.1	65.3	79.0	N/A	81.2
(8)	Zhaohan[30]	87	80	68	N/A	78	N/A	N/A

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

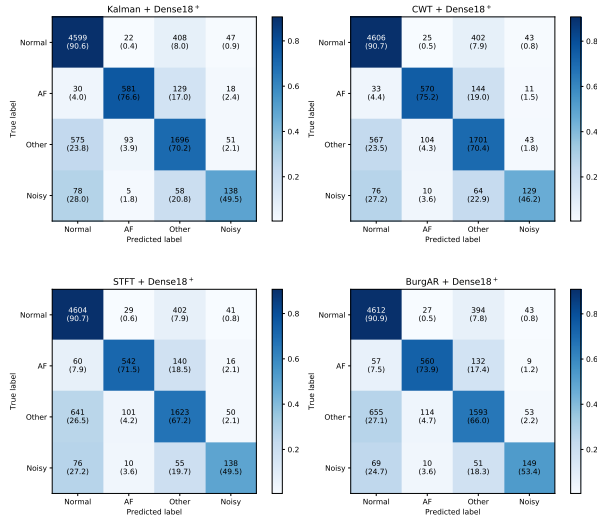**Fig. 6**. Normalized confusion matrix on different methods.

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.

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
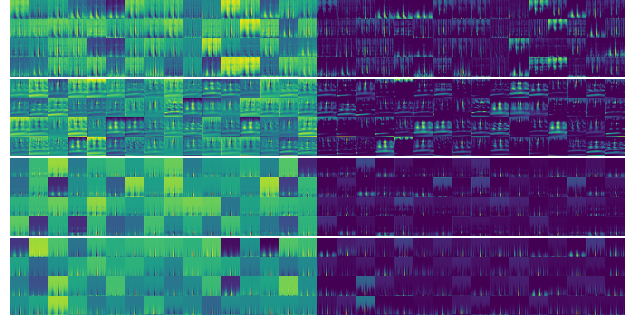
from Other and Noisy classes, which causes low performance in Other and Noisy classes. The classes of the Kalman approach seem have less overlap compared to CWT, STFT, and BurgAR.

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 ($F1_{overall}$) 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
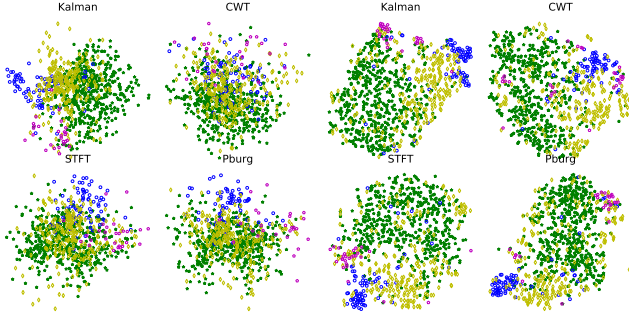
Fig. 8. Left and right 4 groups show VAE and t-SNE visualization on the last concatenate layer of Dense18$^+$ respectively. Data are selected from cross validation test dataset. Green, yellow, blue, and purple represent Normal, Others, AF, and Noise, respectively.

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/smoother 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

[1] S. S. Chugh *et al.*, "Worldwide epidemiology of atrial fibrillation: a global burden of disease 2010 study," *Circulation*, vol. 129, pp. 837–847, 2014.

[2] M. Zoni-Berisso, F. Lercari, T. Carazza, and S. Domenicucci, "Epidemiology of atrial fibrillation: European perspective," *Clinical Epidemiology*, vol. 6, pp. 213–220, 2014.

[3] M. Thaler, *The only EKG book you'll ever need*. Lippincott Williams & Wilkins, 2017.

[4] P. Kirchhof *et al.*, "2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS," *EP Europace*, vol. 18, no. 11, pp. 1609–1678, 2016.

[5] C. Bruser, J. Diesel, M. D. Zink, S. Winter, P. Schauerte, and S. Leonhardt, "Automatic detection of atrial fibrillation in cardiac vibration signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, pp. 162–171, 2013.

[6] F. Yaghouby, A. Ayatollahi, R. Bahramali, M. Yaghouby, and A. H. Alavi, "Towards automatic detection of atrial fibrillation: A hybrid computational approach," *Computers in Biology and Medicine*, vol. 40, no. 11-12, pp. 919–930, 2010.

[7] S. Asgari, A. Mehrnia, and M. Moussavi, "Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine," *Computers in Biology and Medicine*, vol. 60, pp. 132–142, 2015.

[8] M. Mohebbi and H. Ghassemian, "Detection of atrial fibrillation episodes using SVM," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2008, pp. 177–180.

[9] M. Zabihi, A. B. Rad *et al.*, "Detection of atrial fibrillation in ECG hand-held devices using a random forest classifier," vol. 44, 2017.

[10] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT Press, 2016.

[11] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017.

[12] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S. Nemati, "A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology," in *2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*. IEEE, 2017, pp. 141–144.

[13] Y. Xia, N. Wulan, K. Wang, and H. Zhang, "Detecting atrial fibrillation by deep convolutional neural networks," *Computers in Biology and Medicine*, vol. 93, pp. 84–92, 2018.

[14] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–10, 2017.

[15] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," *2017 Computing in Cardiology (CinC)*, vol. 44, 2017.

[16] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[17] G. D. Clifford *et al.*, "AF classification from a short single lead ECG recording: the Physionet/Computing in cardiology challenge 2017," *2017 Computing in Cardiology (CinC)*, vol. 44, 2017.

[18] A. B. Rad *et al.*, "ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 10, pp. 2411–2418, 2017.

[19] A. B. Rad and T. Virtanen, "Phase spectrum prediction of audio signals," in *2012 5th International Symposium on Communications, Control and Signal Processing*. IEEE, 2012, pp. 1–5.

[20] M. Ehrendorfer, *Spectral numerical weather prediction models*. Society for Industrial and Applied Mathematics, 2011.

[21] Y. Qi, T. P. Minka, and R. W. Picara, "Bayesian spectrum estimation of unevenly sampled nonstationary data," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2. IEEE, 2002, pp. II–1473–II–1476.

[22] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013.

[23] J. Hartikainen and S. Särkkä, "Kalman filtering and smoothing solutions to temporal Gaussian process regression models," in *2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2010, pp. 379–384.

[24] S. Särkkä, A. Solin, and J. Hartikainen, "Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 51–61, 2013.

[25] S. Särkkä, "Recursive Bayesian inference on stochastic differential equations," Doctoral dissertation, Helsinki University of Technology, Department of Electrical and Communications Engineering, 2006.

[26] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.

[27] S. M. Kay and S. L. Marple, "Spectrum analysis – a modern perspective," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.

[28] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, vol. 9, 2010, pp. 249–256.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[30] Z. Xiong, M. K. Stiles, and J. Zhao, "Robust ECG signal classification for detection of atrial fibrillation using a novel neural network," *2017 Computing in Cardiology (CinC)*, vol. 44, 2017.

[31] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *13th European Conference on Computer Vision (ECCV)*. Springer, 2014, pp. 818–833.

[32] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, 2014.

[33] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.