Computer Methods and Programs in Biomedicine 221 (2022) 106899

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Deepaware: A hybrid deep learning and context-aware heuristics-based model for atrial fibrillation detection

Devender Kumar[a,*], Abdolrahman Peimankar[b], Kamal Sharma[c], Helena Domínguez[d], Sadasivan Puthusserypady[a], Jakob E. Bardram[a]

[a] Department of Health Technology, Technical University of Denmark, Kgs. Lyngby 2800, Denmark
[b] SDU Health Informatics and Technology, The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense 5230, Denmark
[c] U. N. Mehta Institute of Cardiology and Research Centre, Civil Hospital Campus, Ahmedabad, Gujarat, India
[d] Bispebjerg Hospital, Department of Cardiology, Copenhagen, and Department of Biomedical Sciences at the University of Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 18 November 2021
Revised 20 April 2022
Accepted 17 May 2022

Keywords:
Atrial fibrillation
Deep learning
Electrocardiogram (ECG)
Arrhythmia
Health informatics
Long short-term memory (LSTM)
Context-awareness
Convolutional neural networks

ABSTRACT

Background: State-of-the-art automatic atrial fibrillation (AF) detection models trained on RR-interval (RRI) features generally produce high performance on standard benchmark electrocardiogram (ECG) AF datasets. These models, however, result in a significantly high false positive rates (FPRs) when applied on ECG data collected under free-living ambulatory conditions and in the presence of non-AF arrhythmias.

Method: This paper proposes *DeepAware*, a novel hybrid model combining deep learning (DL) and context-aware heuristics (CAH), which reduces the FPR effectively and improves the AF detection performance on participant-operated ambulatory ECG from free-living conditions. It exploits the RRI and P-wave features, as well as the contextual features from the ambulatory ECG.

Results: DeepAware is shown to be very generalizable and superior to the state-of-the-art models when applied on unseen benchmark ECG AF datasets. Most importantly, the model is able to detect AF efficiently when applied on participant-operated ambulatory ECG recordings from free-living conditions and has achieved a sensitivity (Se), specificity (Sp), and accuracy (Acc) of 97.94%, 98.39%, 98.06%, respectively. Results also demonstrate the effect of atrial activity analysis (via P-waves detection) and CAH in reducing the FPR over the RRI features-based AF detection model.

Conclusions: The proposed DeepAware model can substantially reduce the physician's workload of manually reviewing the false positives (FPs) and facilitate long-term ambulatory monitoring for early detection of AF.

1. Introduction

Atrial fibrillation (AF) is one of the most prevalent cardiac arrhythmias, that is considered as a leading cause of stroke and other heart-related complications in elderly population [6,21]. Nearly 2.3 million people in the USA alone are affected by AF, and this number is likely to increase by 2.5 times by the year 2050 [21]. Early diagnosis and anti-coagulation medication can help in preventing AF complications [61] and ECG analysis is one of the most inexpensive and non-invasive ways for early detection of AF. However, due to its abrupt and paroxysmal nature, it is challenging to detect AF during infrequent and short-term in-hospital checkups. Therefore, there is a great need for enabling longitudinal ambulatory screening and monitoring as a part of the patient's everyday life outside the clinic. Moreover, since visual examination is the usual way for cardiologists to analyze ECG recordings, it is sometimes cumbersome to analyze the huge amounts of data, which would be the result of longitudinal ambulatory ECG recordings. Therefore, in order to realize the ambition of ambulatory cardiac monitoring, it is essential to develop reliable methods for analyzing and interpreting ECG signals and to detect cardiac arrhythmias such as AF.

Over the past two decades, several algorithms have been introduced, which can automatically detect AF from ECG recordings [14,59]. Most of these algorithms are based on classical machine learning and feature engineering techniques (e.g., temporal intervals, wavelet transform, etc.) [32,33,42,56,70]. Feature engineering is an essential step in these models to transform raw data

* Corresponding author.
E-mail addresses: deku@dtu.dk (D. Kumar), abpe@mmmi.sdu.dk (A. Peimankar), kamalcardiodoc@gmail.com (K. Sharma), sapu@dtu.dk (S. Puthusserypady), jakba@dtu.dk (J.E. Bardram).

https://doi.org/10.1016/j.cmpb.2022.106899

into a suitable representation as inputs for the machine learning model to distinguish between different cardiac arrhythmias. Even though feature engineering-based algorithms perform very well in some cases, they face three main challenges: (1) they require hand-crafted feature extraction by a domain expert, (2) they are susceptible to noise in ambulatory settings, and (3) they have relatively low generalization on new data [2,14,43,70].

In recent years, there have been several breakthroughs in the application of DL in areas such as computer vision, natural language processing, and health informatics [10,17,62,65,76,77]. DL has also been widely explored to analyze ECG signals to detect AF in heart disease patients [51]. A number of end-to-end DL models have been introduced for AF detection, which basically bypass the handcrafted feature engineering step needed by other machine learning methods [2,8,16,18,19,24,39,40,47,67,68,72,74]. For example, Wang [68] has proposed a convolutional neural network (CNN) and a modified Elman neural network (MENN) AF detection model, which achieved an accuracy of 97.4% on the MIT-BIH AF Database (AFDB) dataset. Similarly, Faust et al. [19] applied a long-short term memory (LSTM) model for AF detection on heart rate features, which achieved 98.51% accuracy using 10-fold cross-validation on the AFDB dataset. Petmezasa et al. [58] built a hybrid CNN-LSTM model that utilizes focal loss when dealing with an imbalance training set. This model also performs well on the AFDB dataset with a high sensitivity and specificity of 97.87% and 99.29%, respectively.

Despite the promising performance of the above-mentioned research on the publicly available datasets, applying them for longitudinal AF screening in ambulatory free-living conditions still remains an open challenge for several reasons [14,15].

Firstly, most of the DL models have been built and evaluated on public databases, which primarily are short-term recording done in a clinical setting, using high-end clinical-grade ECG recording devices, and which contain manually corrected and annotated R-peak labels [20,54]. The AF detection algorithms based on RRI features are limited by their assumption of receiving almost perfect R-peak detection [54]. However, in contrast to such 'perfect' clinical recordings, ambulatory ECG recordings are often confounded with various artifacts and noise that mimic AF [41] and is often recorded on smaller wearable ECG recording devices with much fewer channels. The presence of noise and low-quality signals makes the detection of R-peaks and P-waves very challenging, if not impossible in some cases. Consequently, this results in non-trivial false positives and performance degradation for such models [16,71]. For instance, in a previous study, we have shown that in a model trained on RRI features, the FPR increased from 1.7% to 4.5% when validating the model on an ambulatory database, which only contains ambulatory normal sinus rhythm (NSR) data from healthy individuals [2]. Moreover, AF episodes occurrences will be rare, especially in the low AF burden population, and noisy ambulatory recordings can often mimic such events. In another study, we have shown that the same AF detection DL model trained on the AFDB dataset achieved excellent performance of around 98% accuracy [31]. However, it resulted in a larger number of non-trivial false positives when applied on patient-operated ambulatory single-channel ECG obtained under free-living conditions. In the study, we found that nearly 62% of all the false positive cases correlated with the participant's ambulatory context under free-living conditions. In particular, they were associated with three types of contexts: (1) change in activity, (2) change in body position, and (3) sudden movement acceleration. It has been shown that incorrect detection of AF in longitudinal screening period could lead to over-diagnosis and patient anxiety [9].

Secondly, to reduce the complexity and achieve real-time detection, most of the AF classification models are primarily trained on the RRI-based features without atrial activity analysis. Such models result in a higher FPR in the presence of non-AF arrhythmias such as premature ventricular contractions (PVCs) and confounding noise in the ambulatory settings, which also exhibit irregular RRI characteristics similar to AF [2,12,66]. In a recent study, Tuboly et al. [66] also highlighted this problem. They showed that in the presence of non-AF arrhythmias, the numbers of false-positive AF detections were significantly higher if relying only on RRI features. Furthermore, Oster et al. [54] also pointed out that AF detection models trained only on RRI (heart's ventricular response) based features are bound to result in high FPR on ECG from free-living conditions [54]. Jalali et al. [26] have tried to address the problem of AF misclassification due to the presence of premature atrial complexes (PACs) by using sensitivity and orthogonality constraints on a Residual Network's cost function. They focused on detecting the irregularities before the AF onset that can indicate the onset of AF. Although this approach showed superior performance in the presence of PACs, the generality of such a model remains unexplored, if it is used in ambulatory conditions with confounding noise and other artifacts.

To reduce the FPR and to improve AF detection on ambulatory ECG recordings with the presence of confounding non-AF arrhythmias, this paper proposes the *DeepAware* model. *DeepAware* is a hybrid multi-fusion and end-to-end AF detection model. The model combines two of our previous algorithms as sub-model [2,57] and combines them with a new context-aware heuristics (CAH) that analyzes and includes the patients' ambulatory contextual data into the model. The context-aware heuristics (CAH)-part of the model specifically enhances the RRI featured based AF detection model's results under the free-living ambulatory conditions. The model is trained using both atrial and ventricular activity types of features.

This paper presents a validation study of the *DeepAware* model. First, we evaluate the model's performance when applied on several existing public datasets and it is shown that the proposed model performs at par, or even better, than existing models. Second, the model's capability of reducing the number of false positive cases, both in the presence of many confounding non-AF arrhythmias as well as under free-living ambulatory conditions, is investigated. *DeepAware* model shows promising results in automatic analysis of longitudinal ambulatory AF screening under free-living conditions. The following are the main contributions of this work:

1. Analysis of the AF detection performance with and without atrial activity features (p-wave detection).
2. Combined deep learning model with context-aware heuristics on ECG from free-living ambulatory conditions that reduces the false-positive rate in RRIs based AF detection model.
3. A highly generalizable model as demonstrated by its performance on 5 different ECG datasets, including two patient-operated datasets from free-living conditions.

The remainder of this paper consists of 5 sections. Section 2 provides the methodology of the proposed algorithm. In Section 3, the proposed *DeepAware* model is described in details. The results are presented and discussed in Section 4. Section 5 presents the limitations and future work, followed by the conclusion in Section 6.

2. Materials and methods

2.1. Databases

In this study, six databases are used to train and validate the performance of *DeepAware*. These include four PhysioNet databases (MIT-BIH AF Database (AFDB) [44], QT database (QTDB) [34], MIT-BIH Arrhythmia Database (MITDB) [45], MIT-BIH Normal Sinus Rhythm Database (NSRDB) [22]) and two *in house*

Table 1
Technical specifications of databases: Ch: No. of ECG channels, Freq: Sampling frequency, NS: Number of subjects in the recording, TR: Total number of records.

Database	Ch	Freq (Hz)	TR	Single Record Length	Total Duration	AF Duration in Hours (%)	Unique Rhythms	NS	Contextual Data
AFDB	2	250	23	10h	234.3h	93.40 (39.87%)	4	25	✗
MITDB	2	360	48	0.5h	24.07h	2.16 (8.97%)	15	47	✗
NSRDB	2	128	18	24h	437.5h	0 (0%)	1	18	✗
CACHET-CADB	1	1024	1602	10sec	4.45h	2.07 (46.6%)	4	24	√
CACHET-NSRDB	1	1024	10	24h	240 h	0 (0%)	1	10	√
QTDB	2	250	105	15min	26.25h	n/a	n/a	105	✗

databases (CACHET Contextualised Arrhythmia Database (CACHET-CADB) [30] and CACHET Normal Sinus Rhythm Database (CACHET-NSRDB)). Technical specifications of these six databases are provided in Table 1.

The QT database (QTDB) contains 105 recordings of 15 minutes each with a sampling rate of 250 Hz, and the annotations include onset, peak, and offset labels of P, QRS, T, and U waves [34]. The AFDB includes 25 long-term ECG recordings of subjects with paroxysmal AF. Among them, two records (#00735 and #03665) were omitted as there is no ECG signals file in the database. For the remaining 23 records, each recording is nearly 10 hours long and has two channels of ECG collected at a sampling rate of 250 Hz [22,44]. The MITDB is sampled at 360 Hz and comprises 48 ECG records of 30 minutes long from 47 subjects. Its annotations files include 15 different rhythms classes [22,45]. On the other hand, MIT-BIH Normal Sinus Rhythm Database (NSRDB) [22] contains 18 long-term two channels ECG (sampled at 128 Hz) from healthy subjects, which are mostly in NSR without any significant arrhythmias.

The two in-house databases (CACHET Contextualised Arrhythmia Database (CACHET-CADB) and CACHET Normal Sinus Rhythm Database (CACHET-NSRDB)) are collected as part of the mCardia's feasibility study [28,29] conducted in Denmark and India. Due to its technical nature, this study is exempted from ethical approval both by the Danish National Committee on Health Research Ethics (File #H-19071015) and by the Institutional Review Board (IRB) of Mahatma Gandhi University of Medical Sciences and Technology, Jaipur India. The single channel Movisens EcgMove4 ECG monitor [48] is used for collecting both databases. In addition to the ECG sensor (1024 Hz), the EcgMove4 device also contains a 3D acceleration sensor (64 Hz), rotation rate sensor (64 Hz), and pressure sensor (8 Hz). The participants are recruited during their outpatient arrhythmia clinic visits. Preference is given to the participants who were either already diagnosed with AF or at a high risk of AF. It is also ensured that they have an active life and are not bed-ridden or critically ill. The average age of the participants is 59 years. Participants continuously wear the chest-mounted EcgMove4 device. The recording length in CACHET-CADB vary from a single day to over two weeks. The CACHET-NSRDB is collected from healthy individuals using the same hardware and software setup as in CACHET-CADB.

Thus, both the CACHET-CADB and CACHET-NSRDB contain participant operated single-channel contextualised ECG obtained under free-living conditions. Beside the ECG recordings, the participant's ambulatory contextual information such as activities, body positions, and movement accelerations are also recorded. Contextual data is obtained by processing the raw data from the accelerometer, rotation rate sensor, and pressure sensor of the chest-mounted ECG device for every 10 seconds interval using the Movisens DataAnalyzer tool [49,50]. These activities are aggregated every 10-seconds and contain activities such as lying, sitting/standing, cycling, jogging, and walking. A while-box decision tree is used to calculate these activities from the combination of features derived from the 3D accelerometer and the barometric air pressure sensor data [49]. In addition, body positions such as lying supine/left/right/prone, upright, sitting, and standing are derived using the inclination obtained from the raw acceleration signals [49].

There are 1602 manually annotated ECG records of 10 seconds long from 24 subjects in the CACHET-CADB. Each record belongs to one of the four classes, namely 'AF', 'NSR', 'noise', and 'others'. The manual annotations of these ECG records are done by two cardiologists independently, and it only includes labels with 100% inter-rater agreements between them. The CACHET-NSRDB contains 10 long-term NSR ECG records from healthy individuals. All the recordings in CACHET-NSRDB are almost over 24 hours long. Please note that the hardware, contextual information collection and processing steps remain the same in both CACHET-CADB and CACHET-NSRDB. These two datasets are used in the context-aware heuristics (Section 3.5) under free-living conditions as they provide the participant's ambulatory contexts information during the ECG recording periods.

2.2. Deep learning

DL enables computational models to learn useful features directly from the input data [37]. It has enhanced the state-of-the-art algorithms in domains such as image and speech recognition, natural language processing, drug discovery, and genomics [37,65]. In recent years, DL has successfully been applied for the detection of AF and other types of arrhythmias [1,16,35].

2.2.1. CNN layer

CNN [38,53] have been proven very efficient in pattern recognition tasks by exploiting both the spatial and temporal patterns in the data [37]. To achieve this, CNNs follow four key steps: 1) local connections; 2) shared weights for convolution process; 3) create large number of filters; and 4) reduce the network complexity as much as possible. Besides the input and output layers, a typical CNN structure consists of one or more connected convolutional layers, pooling layers, ReLU, and normalization layers. Fig. 1 depicts a typical CNN structure. In 1D-CNNs for analyzing ECG signals, various filters are generated by sliding a fixed window over the ECG record. The size of the window is known as the kernel size (k_{size}). The weights of these kernels and the overall bias is to be learned during the training process. It should be noted that the weights of the kernel are fixed for each filter map [23].

2.2.2. LSTM layer

Recurrent Neural Networks (RNNs) are specially designed to efficiently capture dependencies in sequential information within time-series data. However, it has been shown that learning long-term dependencies are very challenging [3]. On the other hand, the problem of unstable gradient can be solved by LSTM networks (special type of RNNs), which can handle long-term dependencies [25]. As shown in Fig. 2, a LSTM block has three main parts: 1) input gate (i_t), 2) forget gate (f_t), and 3) output gate (o_t). Forget and input gates control the flow of information removal and addition to the memory block as follows:

$$f_t = \sigma\left(\mathbf{u}_f^T \mathbf{a}_t + \mathbf{w}_f^T \mathbf{h}_{t-1} + b_t\right), \quad (1)$$

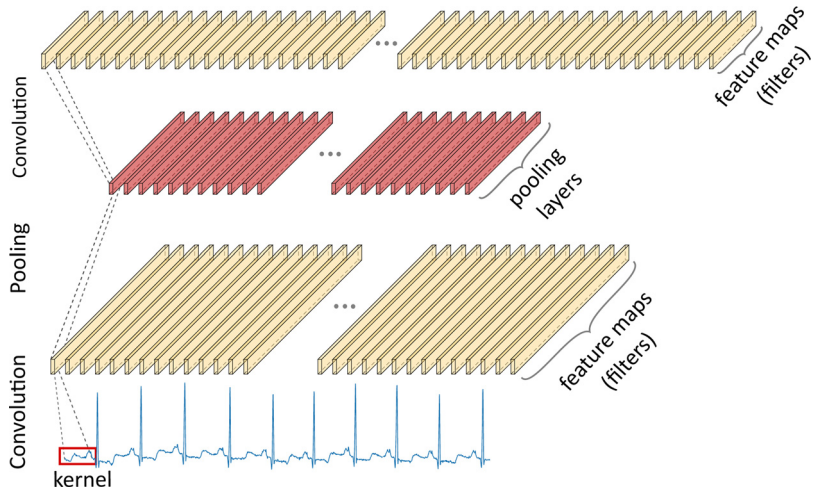

Fig. 1. Typical CNN structure.

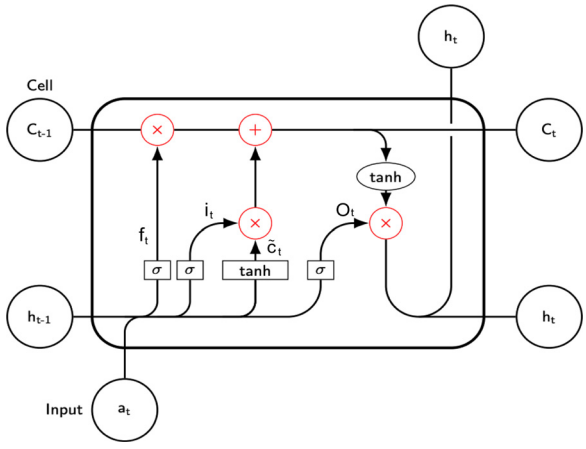

Fig. 2. LSTM memory block.

$$i_t = \sigma(\mathbf{u}_i^T \mathbf{a}_t + \mathbf{w}_i^T \mathbf{h}_{t-1} + b_t), \tag{2}$$

where $\mathbf{a}_t$ is the output from the previous layer and is the input to the LSTM block at time step t, and $\mathbf{h}_{t-1}$ is the output of the LSTM block at time $t-1$. The trainable parameters of the LSTM block are $\mathbf{w}_f$, $\mathbf{u}_f$, $\mathbf{w}_i$, $\mathbf{u}_i$, b_f, and b_i, which are weight vectors and bias terms. The memory of a LSTM block, c_t, is updated as follows:

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t, \tag{3}$$

where, $\tilde{c}_t = \tanh(b_c + \mathbf{u}_c^T \mathbf{a}_t + \mathbf{w}_c^T \mathbf{h}_{t-1})$. Consequently, the output of the LSTM block is generated by:

$$h_t = o_t \tanh(c_t), \tag{4}$$

where $o_t = \sigma(\mathbf{w}_o^T \mathbf{a}_t + \mathbf{u}_o^T \mathbf{h}_{t-1} + b_o)$. Here, $\mathbf{u}_o$ and $\mathbf{w}_o$ are the weight vectors and b_o is the bias of the output gate. This means that the LSTM is capable of keeping or forgetting the existing memory efficiently [11].

Bidirectional LSTM (BiLSTM) is a variant of LSTM, which unlike LSTM can process the sequential time-series in both forward and backward directions with two separate hidden layers. BiLSTM have been found very useful in several ECG classification algorithms [2,75]. The *DeepAware* algorithm proposed in this study combines two DL models from our earlier studies [2,57]. The first model (denoted as the RR-Net in Fig. 3) is a combination of CNN and BiLSTM layers, which takes the RRIs as inputs [2,63]. The second model, named as the DENS-ECG [57] in Fig. 3, is a combination of CNN and BiLSTM layers, which is used for the P-wave detection from ECG. As shown in Fig. 3, there are three layers of CNN followed by one BiLSTM layer. DENS-ECG takes raw ECG signals and outputs the number of detected P-waves. The output of these two models is combined with the context-aware heuristic model to detect the AF rhythms.

3. DeepAware architecture

Fig. 3 illustrates the flowchart of the proposed *DeepAware* algorithm. It comprises of six main components: (1) ECG data preprocessing, (2) segmentation, (3) RR-Net, (4) DENS-ECG, (5) the context-aware heuristic model, and (6) the AF decision box. In this section, all these six components are described in detail.

3.1. Data preparation and pre-processing

As shown in Fig. 3, data preparation and pre-processing is the first step. The first channel of the ECG records of PhysioNet datasets has been used and all the six datasets (Table 1) are resampled to 250 Hz. It should be noted that only a single channel ECG of the PhysioNet dataset was used because two of our in-house datasets, CACHET-CADB and CACHET-NSRDB (see Table 1), have only a single channel. Also, the first channel (channel 1) out of the two channels in the PhysioNet datasets was used as it gave us slightly better performance in the initial experiments. The baseline wanders (<0.5Hz) and high-frequency noises (>40Hz) are removed using a band-pass (0.5–40 Hz) filter. The ECG signals are then smoothed by a Savitsky-Golay filter [60]. The Savitsky-Golay filter effectively smoothens the signal and increases precision without distorting the signal tendency (which helps in improving P-wave detection in the ECG). It should be noted that, in some databases like AFDB, MIT-BIH Arrhythmia Database (MITDB), and NSRDB, the R-peak locations are already available within the database, whereas, for CACHET-CADB and CACHET-NSRDB, the Pan-Tompkins algorithm [55] is used for finding the R-peaks locations. Since the CACHET-NSRDB contains continuous ambulatory ECG data and therefore high levels of noise, a cross-correlation (auto-correlation as they are single channel) based noise detector is applied prior to the Pan-Tompkins algorithm. The ECG singles are segmented in the sliding windows of 10-second, and the windows with the cross-correlation value <0.65 are rejected as noise and timestamped. This correlation cutoff value is chosen through repeated experimentation to ensure that we do not reject the "good signal" in the preprocessing step, even if it means

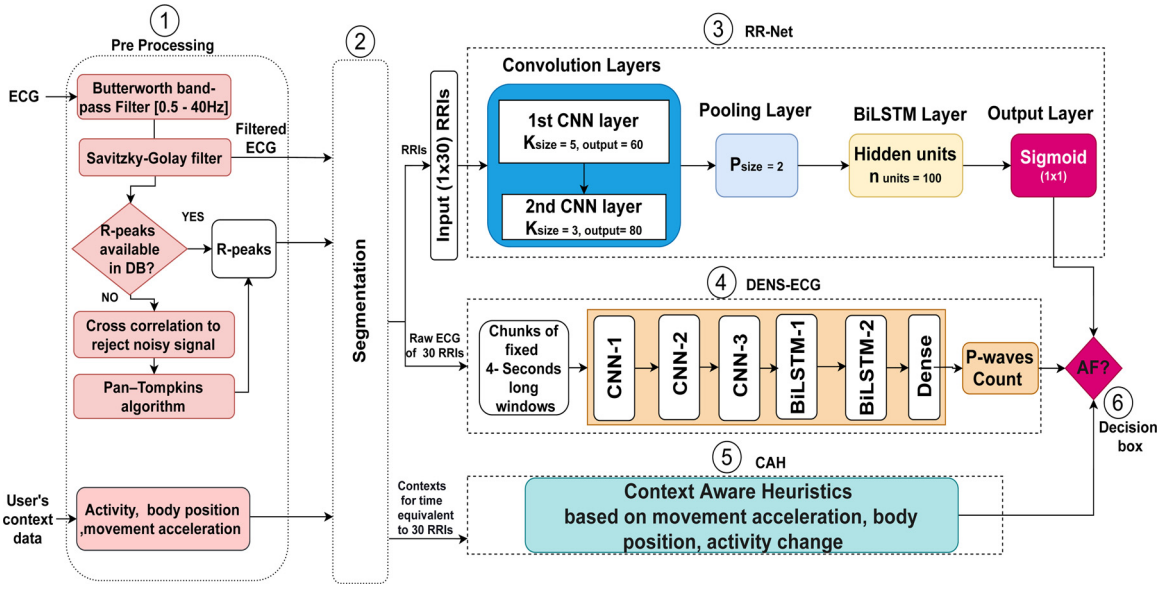

Fig. 3. The architecture of the proposed *DeepAware* model consists of six sub-components: (1) ECG data preprocessing, (2) segmentation, (3) the RR-Net, which take inputs of the RR interval series, (4) the *DENS-ECG*, which takes the raw ECG inputs and gives P-wave count, (5) and a CAH model, which takes user context in a case of ambulatory ECG to check if any change in user's context is detected, and (6) AF decision box for final binary output.

allowing some level of noise to pass. In this process, nearly 6% of the CACHET-NSRDB's 10-second segments were labelled as noise and rejected. To check the Pan-Tompkins algorithm's accuracy for R-peaks detection on the CACHET-CADB and CACHET-NSRDB, we visually inspected some (nearly 0.01%) random ECG samples of its output on these databases. No samples were rejected in this process. It should be mentioned that the user's ambulatory context data (i.e., activity, body position, movement acceleration) in the CACHET-CADB and CACHET-NSRDB is already available for every 10 seconds of intervals and is used without any further processing.

3.2. Segmentation

The RRIs and the filtered ECGs are segmented into a window length of 30 RRIs. The sliding window has an overlap of 10 RRIs. The segmented windows are provided as inputs to the RR-Net model. Please note that we have also experimented with various window lengths (i.e., 10, 15, 20, 25, 30, 35, 40) and overlapping for the RR-Net sub-module. Similar to Andersen et al. [2], the 30 RRIs window with an overlap of 10 RRIs gave us the best results on AFDB and MITDB. Therefore, for the final experiment, the input window length of 30 RRIs is chosen. The corresponding ECG segments, which has the same size as 30 RRIs, is fed simultaneously as inputs to the DENS-ECG model. As shown in Fig. 3, this ECG is further lumped into fix windows of 4 seconds length before it is passed to the input layer of DENS-ECG.

Similarly, the CACHET-CADB and CACHET-NSRDB databases, which include the information about the user's context are also segmented into time-duration equal to that of 30 RRIs. The 10-seconds segments are combined before the RRI calculation in CACHET-CADB, and thereafter the segmentation and windowing process is performed.

3.3. RR-Net

Irregular RRIs is considered as one of the strong indications of AF. The RR-Net model is a combination of two convolutional layer followed by a BiLSTM layer [2]. The convolutional layers extract the features from the RRIs, which are used by the BiLSTM layer afterwards. The first convolutional layer uses a kernel of size 5 ($K_{size} = 5$) and outputs 60 features. The input sequences are zero-padded to preserve the temporal dimension. The second convolution layer has a K_{size} of 3 and generates more abstract features. Here again, zero-padding is applied to preserve the temporal dimension. As depicted in Fig. 3, a max pooling layer is applied after the two convolutional layers, which has a kernel size (P_{size}) of 2 with strides of two. This layer results in reducing the temporal dimension of the inputs by half, which is an essential step for bringing down the complexity before the BiLSTM layer. The output of the pooling layer is fed into the BiLSTM layer consisting $n_{units} = 100$ hidden units. The output of the BiLSTM is fed into the classification layer with a sigmoid activation. The output can be considered as the posterior probability of the degree of irregularity for the i^{th} RRIs (input sequence). These probabilities are finally converted to a binary output of the RR-Net model as follows:

$$\text{RR-Net}(i) = \begin{cases} 1, & \text{if } p(y_i = irregular | \mathbf{x}_i, \text{RR-Net}) \geq 0.5, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where RR-Net(i) is the binary output of the i^{th} RRI segment in which 1 represents AF and 0 represents Non-AF. $p(y_i = irregular | \mathbf{x}_i, \text{RR-Net})$ is the probability output of the sigmoid function of the RR-Net for the i^{th} RRI segment. It should be noted that the probability threshold is set to 0.5.

3.4. DENS-ECG model

The DENS-ECG model is a combination of three convolutional layers and a dropout layer followed by two BiLSTM layers [57]. It takes 4 seconds long windows of raw ECG segments as inputs to the first 1D convolution layer to delineate the ECG signals. The 1D convolutional layers extract abstract features from ECG segments. The two BiLSTM layers are used to process the extracted features by the previous 1D convolutional layers. The three convolutional layers use a kernel size of 3 and the number of filters (feature maps) for the three successive layers are 32, 64, and 128, respectively. In addition, zero padding is applied to maintain the same dimension in the input and convolutional layers. For example, the output of the third convolutional layer is 128 feature maps, which are then used as inputs for the first BiLSTM layer. The corresponding number of hidden units (n_{units}) are 250 and 125 for the two

BiLSTM layers, respectively. Finally, The output of the second BiLSTM layer is fed into a dense layer, which generate posterior probabilities for the P-, QRS, T-, and No-wave segments of the ECG signals. As presented in Fig. 3, the number of P-waves detected by DENS-ECG model is provided to the decision box. It is worth noting that the dropout layer after the third convolutional layer helps in preventing the over-fitting problem during the training phase of the model. The dropout probability is set to 0.2, which means that 20% of the units is set to zero at each training step. The absence of P-waves for the i^{th} ECG segment is computed as follows:

$$\text{DENS-ECG}(i) = \begin{cases} 1, & \text{If } P_c \leq 15 \text{ for 31 R peaks (30 RRIs)}, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where DENS-ECG(i) is the predicted P-wave for the i^{th} RRIs segment and P_c is the number of P-waves detected by the DENS-ECG model for the i^{th} ECG segment as the threshold is set to 15.

3.5. Context-aware heuristics

The context-aware heuristics (CAH) model is based on our previous work [31] in which we analyzed the relationship between the FPR and user's context on an AF detection model trained on RRI features. The analysis of false-positive cases using contextual data concluded that the vast majority (∼78%) of short (<50 Seconds) FP segments were associated with three main contexts: 1) change of activity; 2) change in body position (especially during laying/sleep); and 3) sudden movement acceleration.

The CAH takes activity, body position, and movement acceleration as input and evaluates if there is a change in the user's context during a specific 30 RRIs segment or its preceding segments. As shown in Eqn. (7), the CAH model assigns a binary output to detect whether a context change (i.e. change in activity, change in body position or sudden movement acceleration) is detected during the current or previous RRI input windows. Any identified context changes resulting in a non-AF episode detection for the corresponding RRIs segment are specified as follows:

$$\text{CAH}(i) = \begin{cases} 0, & \text{if context change detected}, \\ 1, & \text{otherwise}, \end{cases} \tag{7}$$

where CAH(i) is the prediction for the i^{th} RRIs segment.

3.6. The decision box

As depicted in Fig. 3, the outputs of these three models (RR-Net, DENS-ECG, and Context-Aware Heuristics) are combined at the decision box. This is performed as follows:

$$\widehat{D}(i) = \begin{cases} \text{RR-Net}(i) \wedge \text{DENS-ECG}(i), \text{If CAH unavailable}, \\ \text{RR-Net}(i) \wedge \text{DENS-ECG}(i) \wedge \text{CAH}(i), \text{otherwise}, \end{cases} \tag{8}$$

where $\widehat{D}(i)$ is the final binary classification for the i^{th} input sequence. The operator $\wedge$ is the logical "and", which combines the output of sub-modules (RR-Net(i), DENS-ECG(i), and CAH(i)).

3.7. Model training

The AFDB and QTDB datasets are used for training the RR-Net and DENS-ECG sub-modules, respectively. The DENSE-ECG is trained using stratified 5-fold cross validation technique [57]. Similarly, the RR-Net is trained on AFDB using 10-fold cross validation techniques [5] and the data is split segment-wise. The RR-Net sub-module is trained using the Stochastic Gradient Descent (SGD) with Nesterov accelerated gradient [52], whereas an Adam optimization algorithm [27] is used in the DENS-ECG model. In both RR-Net and DENS-ECG, the hyper-parameters are fine-tuned using a random search technique [4]. The number of trainable parameters in DENS-ECG and RR-Net are 1,416,044 and 159,841, respectively.

It should be noted that the CAH sub-module is tested/evaluated only on CACHET-CADB and CACHET-NSRDB. It takes the activity, body position, and movement acceleration as input and keeps track of any changes in these three contexts in both the current or previous input widows. As mentioned in Sections 3.1 and 3.2, the continuous activity, body position, and movement acceleration are already preprocessed and available in both CACHET-CADB and CACHET-NSRDB. All the individual sub-modules of *DeepAware* (Fig. 3) are built in python 3.7 using the Tensorflow 2.4.1 framework. The entire training process is done on a MacBook Pro running MacOS 10.15.7 with 16 GB RAM, Dual-Core Intel Core i7 processor and an Intel Iris Plus Graphics 650 1536 MB graphics card.

3.8. Statistical analysis

The FPR outputs of the proposed DeepAware model and its variants (i.e., RR-Net $\wedge$ CAH, and RR-Net $\wedge$ DENS-ECG) are compared with RR-Net model on the test datasets using a paired-samples t-test for statistical significance. The paired sample t-test is a statistical method employed to find if the mean difference between two sets of observations is zero. The null hypothesis for the paired-samples t-test is that the means of the FPR in the two models are the same. Comparisons with p-values < 0.05 are considered statistically significant.

4. Results and discussion

The RR-Net and DENS-ECG submodules were first individually trained and tested using cross-validation on AFDB and QTDB, respectively. Then, the best performing models from the cross validation process (i.e. fold 2 for RR-Net and fold 1 for DENS-ECG) were selected. Finally, the best performing models were evaluated on the MITDB, NSRDB, CACHET-CADB, and CACHET-NSRDB, which are unseen datasets to the models.

The MITDB dataset contains 14 types of non-AF arrhythmias and the performance of the model on this dataset indicate its generalizability in the presence of PVCs and other non-AF arrhythmias. On the other hand, both the NSRDB and CACHET-NSRDB datasets only contain normal sinus rhythms, which are used to evaluate the performance of the model and its expected FPR on healthy subjects. In addition, the CACHET-NSRDB and CACHET-CADB datasets contain the user's contextual information during ambulatory ECG recordings under free-living conditions. These two datasets are specifically used in the context-aware heuristics which keeps track of any changes in the user's ambulatory context. The metrics used for evaluating the performance of *DeepAware* and the obtained results on each dataset are described in the following sections.

4.1. Model evaluation metrics

To report the performance of our model, we apply the standard metrics of a confusion matrix, namely, the average accuracy (Acc), sensitivity (Se), specificity (Sp), and FPR, which are defined in Eqs. (9)–(12), respectively. In a confusion matrix (see Table 3), each row matrix represents the instances in an actual class while each column represents the instances in the predicted class.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}. \tag{9}$$

$$Se = \frac{TP}{TP + FN}. \tag{10}$$

Table 2
Comparison of *Deep Aware* algorithm with other state-of-the-art models on AFDB, MITDB, and NSRDB datasets. All the results are in percentage. Ch: Number of ECG channels, MFSWT: Modified Frequency Slice Wavelet Transform, MCNN: Multi-Scale CNN, HAN: Hierarchical Attention Network, MENN: Modified Elman Neural Network, IHRS: Instant Heart Rate Sequence, RCN: Recurrence Complex Network, FFS: F-wave Frequency Spectrum. Note that the most of the published articles on AF detection reported their performances on only AFDB, and their generalizability on MITDB and NSRDB are not reported or non-comparable (as some have combined these datasets). The '-' in the table below implies that comparisons are not available in the published articles.

Algorithm	Methods	Features	Ch	AFDB				MITDB				NSRDB	
				Se	Sp	Acc	FPR	Se	Sp	Acc	FPR	Sp	FPR
[13]	CNN, BLSTM	Heartbeat Sequences, RRI	1	99.93	97.03	96.59	-	-	-	-	-	-	-
[73]	CNN	MFSWT	1	74.96	86.41	81.07	-	-	-	-	-	-	-
[71]	CNN	SWT, STFT	1	98.79	97.87	98.63	-	-	-	-	-	-	-
[36]	CNN	RRI, FFS	1	97.4	96.2	97.3	-	-	-	-	-	-	-
[69]	CNN, RCN	Raw ECG	2	94.28	94.91	94.59	-	-	-	-	-	-	-
[74]	MCNN	IHRS	2	98.22	98.11	98.18	-	-	-	-	-	-	-
[46]	BiRNN, HAN	Raw ECG	2	99.08	98.54	98.81	-	-	-	-	-	-	-
[68]	CNN, MENN	Raw ECG	2	97.9	97.1	97.4	-	-	-	-	-	-	-
[2]	CNN, BiLSTM	RRI	2	98.17	96.29	97.1	3.71	98.96	86.04	87.4	13.96	95.01	4.99
DeepAware	CNN, BiLSTM	Raw ECG, RRI, Context	1	98.27	98.84	98.62	1.16	93.05	91.67	91.82	8.33	98.47	1.53

Table 3
Confusion matrix.

	Predicted positive	Predicted negative
Actual positive	True positive (TP)	False negative (FN)
Actual negative	False positive (FP)	True negative (TN)

$$Sp = \frac{TN}{TN + FP}. \tag{11}$$

$$FPR = \frac{FP}{FP + TN}. \tag{12}$$

4.2. The performance of DeepAware on public datasets and its comparison with the literature.

A direct comparison of the performance deep learning models for AF detection in the literature is challenging due to factors such as different types of data acquisition processes and devices used, kinds of ECG features used, differences in training, and training/testing/validation data splitting (inter-patient or intra-patient) [40]. Despite these experimental differences among studies, Table 2 attempts to compare the performance of the proposed DeepAware model with other state-of-the-art models on the AFDB, MITDB, and NSRDB datasets. Fig. 4a and 4b also show the confusion matrices of the proposed *DeepAware* model on AFDB and MITDB datasets, respectively. The *DeepAware* model clearly outperforms the state-of-the-art algorithms on AFDB and it is generalized enough to perform well on unseen datasets such as MITDB and NSRDB. It achieved a sensitivity, specificity, and accuracy of 98.27%, 98.84%, and 98.62%, respectively, on AFDB dataset using a 10-fold cross-validation.

The proposed model also achieved a sensitivity, specificity and accuracy of 93.05%, 91.67%, and 91.82% on MITDB dataset, respectively. Compared to the results presented in Andersen et al. [2] on the MITDB dataset, *DeepAware* improves the specificity and accuracy by 5.63% and 4.42%, respectively, at the cost of 5.91% reduction in sensitivity. It should be mentioned that the performance of the *DeepAware* model on the MITDB dataset indicates its robustness in the presence of PVC/VPC beats and non-AF arrhythmias. The ectopic beats and non-AF arrhythmia resemble the AF in terms of irregularity in the RR intervals, thereby causing more FPs in AF detection models [7,64,66]. For example, in Andersen et al. [2] (c.f., Table 2), despite high specificity on the AFDB dataset, the model's specificity on the MITDB, which has 14 other types of non-AF arrhythmias, has reduced drastically to 86.04%.

Furthermore, *DeepAware* seems to generalized well and performs better on the NSRDB dataset compared to other state-of-the-art models. As it can be seen in Table 2, the *DeepAware* model improved the specificity reported in Andersen et al. [2] by 3.57%.

4.3. The performance of deepaware on contextualised ECG datasets

The context-aware heuristics are applied on the CACHET-CADB and CACHET-NSRDB, both of which contain contextual data from the ambulatory ECG recordings under free-living conditions. As shown in Table 4, the proposed *DeepAware* model has achieved a sensitivity, specificity, and accuracy of 97.94%, 98.39%, 98.06% on

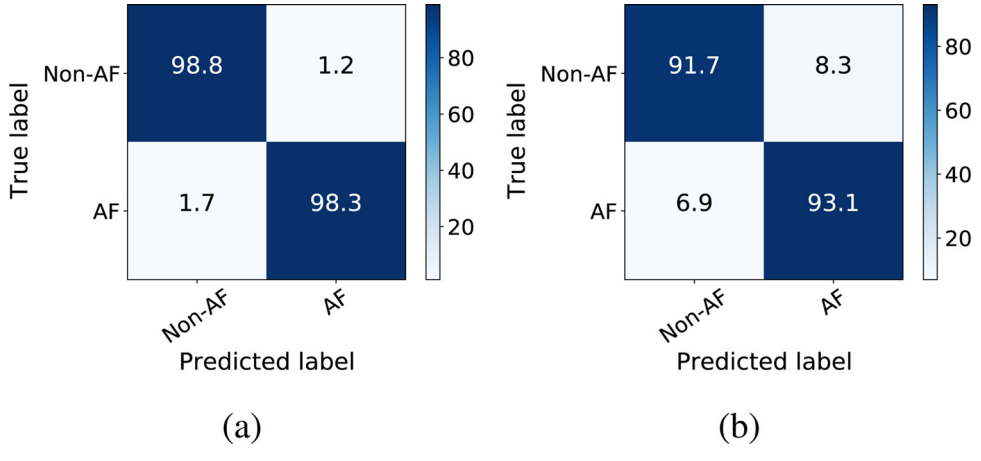

Fig. 4. Confusion matrix of: (a) AFDB and (b) MITDB. The numbers are in percentage.

Table 4
Performance on CACHET-CADB. The RR-Net is a sub-module (Fig. 3) trained on just RRI-features. The comparison is to highlight the improvements made by DeepAware as compared to just relying on RRI features.

Measure	CACHET-CADB			
	RR-Net	RR-Net ∧ CAH	RR-Net ∧ DENS-ECG	DeepAware
Se (%)	99.63	99.44	97.94	97.94
Sp (%)	90.32	94.64	98.39	98.39
Acc (%)	97.22	98.19	98.06	98.06
FPR (%)	09.68	5.38	01.61	01.61

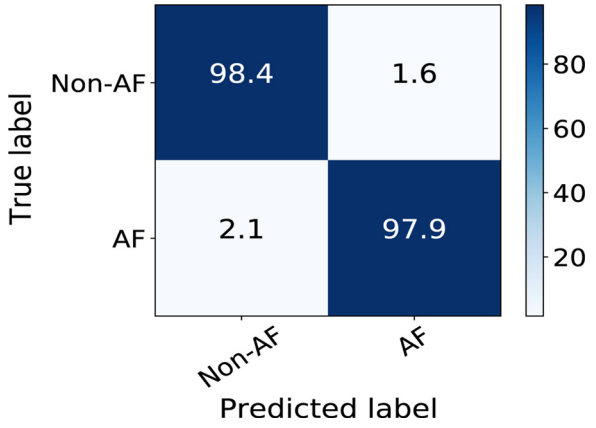

Fig. 5. Confusion matrix of CACHET-CADB. All numbers are in percentage.

the CACHET-CADB dataset, respectively. Fig. 5 shows the confusion matrix of the *DeepAware* model when applied on the CACHET-CADB dataset. Similarly, Table 5 reports the performance of the *DeepAware* model on the CACHET-NSRDB dataset. The average FPR for all the records in Table 5 is 1.76%. As shown in Table 5, in general, the proposed *DeepAware* model outperforms the RR-Net model on the CACHET-NSRDB dataset.

Additionally, as shown in Tables 4 and 5 a comparison between the performance of RR-Net and *DeepAware* on CACHET-CADB and CACHET-NSRDB confirms the positive effect of the context-heuristics and DENSE-ECG to lower the FPR on participants-operated ECG under free-living conditions. We can see that on the CACHET-CADB dataset, *DeepAware* has improved the specificity and reduced the FPR by around 8% at the cost of a 1.69% reduction in the sensitivity. Similarly, on CACHET-NSRDB, the average FPR has been reduced by 7.81%. Furthermore, since CACHET-NSRDB only contains subjects having normal sinus rhythms, the performance of *DeepAware* on CACHET-NSRDB is a good indication of expected FPR in healthy and low AF prevalence subjects under free-living conditions.

4.4. AF Detection with and without atrial activity analysis

Atrial activity analysis is done by checking the existence of a P wave as the P-wave in ECG represents atrial depolarization. In the absence of a P-Wave detection, AF detection models relying on features of only RRIs cannot distinguish between AF and other arrhythmias with irregular RRI (e.g., sinus arrhythmia, premature ventricular contraction) [66]. Tables 4–6 show a comparison between the performances of RR-Net and *DeepAware* (as RR-Net ∧ DENS-ECG) on all the four test datasets. These results show the impact of including atrial activity analysis (i.e., P-wave detection using the DENS-ECG model) on the FPR in the AF detection algorithm. In Table 6, the FPR on MITDB and NSRDB is reduced by 4.57% and 2.94%, respectively. Similarly, on CACHET-CADB (Table 4) and CACHET-NSRDB (Table 5), DeepAware has improved the FPR by 8% and an average of 7.3%, respectively. The paired-samples *t*-test between the output of RR-Net and DeepAware for FPR on all these test datasets has p-value = 0.017, implying the statistically significant reduction in FPR by DeepAware, which is due to the addition of atrial activity (P-wave) analysis.

These results are consistent with the findings by Tuboly et al. [66], which also showed that taking atrial activity analysis features into account can significantly reduce the FPR as compared to just RRI features based AF detection models. It should be noted that respiratory sinus arrhythmia, which is a natural response of the healthy heart, can be misclassified as AF without the analysis of atrial activity. This false diagnosis usually occurs in the young population with a low prevalence of AF [66]. It is also important to highlight that existing literature on deep-learning-based AF detection has limited coverage of examining the impact of non-AF arrhythmias (i.g, sinus arrhythmia) on FPR and specificity of AF detection algorithms.

4.5. Applying context-aware heuristics on RRI-based model

Fig. 6 shows a typical scenario of an ECG signal captured under free-living ambulatory conditions. The irregularity in RRI, which is induced by a change in the user's ambulatory contexts, may lead to an AF diagnosis. Such RRI irregularities caused by changes in the context are either the heart's natural response to change or can be due to motion artifacts. The RRI irregularity induced by a change in the context is usually short (30–60 seconds) [31]. Therefore, it becomes difficult for RRIs based models (such as RR-Net sub-model here) to identify whether the irregularity in RRIs is due to AF or the sudden change in ambulatory contexts. The context-aware heuristics (CAH) in *DeepAware* model helps to identify whether the RRI irregularity detected by RR-Net is, in fact, due to heart disease or it is because of the change in the user's ambulatory contexts.

The impact of combining the RR-Net with context heuristics (i.e., RR-Net ∧ CAH) in reducing the FPR under free-living ambulatory conditions can be observed in both Tables 4 and 5. For CACHET-CADB, the RR-Net ∧ CAH reduces the FPR by nearly 4.3% compred to RR-Net model, whereas for CACHET-NSRDB, it reduces the FPR by 4.6% on average. The paired-samples *t*-test between RR-Net and RR-Net ∧ CAH on CACHET-NSRDB achieves p-value of 0.0003, which shows a statistically significant reduction in the FPR over RR-Net.

It can be seen that using CAH along with RR-Net ∧ DENS-ECG on CACHET-CADB (Table 4) has no significant effect since the performance of RR-Net ∧ DENS-ECG and DeepAware (RR-Net ∧ DENS-ECG ∧ CAH) are the same. It is due to the fact that although the RR-Net detects the RRIs irregularity on context changes and classifies it as AF (which CAH tries to prevent), the detected P-waves in the signal by DENS-ECG has already guaranteed that it is not classified as AF. In addition, it is very likely that CACHET-CADB, as a small dataset, does not contain the cases where the effectiveness of CAH can be seen. The results on CACHET-NSRDB (Table 5) highlight this point as the output of RR-Net ∧ DENS-ECG and DeepAware (RR-Net ∧ DENS-ECG ∧ CAH) are different. It should be also noted that in Table 5, the FPR differences between RR-Net ∧ DENS-ECG and DeepAware are minor for a few records (e.g., record no 4, 9). However, the paired-samples *t*-test between RR-Net ∧ DENS-ECG and DeepAware (RR-Net ∧ DENS-ECG ∧ CAH) for all the ten records has a p-value = 0.01.

These results indicate that CAH can be especially more effective in reducing FPR in models that rely only on RRI features for AF detection. But CAH is less effective when a P-wave detection model, such as DENS-ECG, is sequentially applied before it. Although the P-wave detection model helps reducing FPR on CACHET-CADB, the proposed context context-aware heuristics (CAH) can be specially more useful in the presence of multi-class arrhythmias and more complicated ECG morphologies where P-wave detection is chal-

Table 5
Performance on CACHET-NSRDB. The RR-Net is a sub-module (Fig. 3) trained on just RRI-features. The comparison is to highlight the improvements achieved by DeepAware as compared to just relying on RRI features. Each record consists of over 24 hours long contextualised ECG under free living conditions from healthy individuals. Input No.: Number of (30x1) input windows.

Record	Input No.	R-Peaks	RR-Net		RR-Net ∧ CAH		RR-Net ∧ DENS-ECG		DeepAware	
			Sp	FPR	Sp	FPR	Sp	FPR	Sp	FPR
1	5714	114,319	89.10	10.90	93.73	6.27	97.22	2.78	98.41	1.59
2	5906	118,156	88.52	11.48	93.89	6.11	94.18	5.82	95.55	4.45
3	3998	80,037	89.37	10.63	95.37	4.63	99.30	0.70	99.39	0.61
4	3535	70,733	91.85	08.15	97.17	2.83	99.77	0.23	99.80	0.20
5	1429	28,634	97.06	02.94	98.10	1.90	98.32	1.68	99.02	0.98
6	4123	82,565	82.77	17.23	90.49	9.51	97.79	2.21	98.16	1.84
7	5388	108,046	95.36	04.64	96.25	3.75	96.73	3.27	96.82	3.18
8	5959	119,276	80.87	19.13	89.26	10.74	96.02	3.98	96.29	3.71
9	4600	92,173	94.04	05.96	97.54	2.46	99.41	0.59	99.54	0.46
10	5017	100,396	95.26	04.74	98.33	1.67	98.82	1.18	99.36	0.64

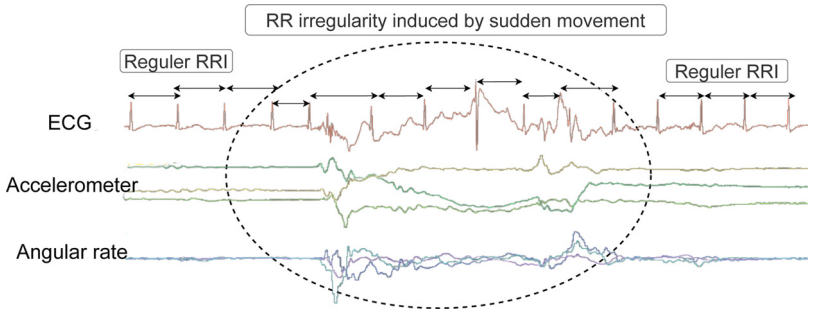

Fig. 6. An example of irregular RRI caused by changes in user's ambulatory context, which resembles an AF episode. The figure shown the single-lead ECG signal (top), accelerometer data (middle), and angular rate (bottom).

Table 6
Classification performance of RR-Net and DeepAware on the MITDB and NSRDB datasets. Note that the RR-Net is a sub-module (Fig. 3) trained on just RRI-features. The comparison is to illustrate the improvements made by DeepAware as compared to just relying on RRI features.

Measure	MITDB		NSRDB	
	RR-Net	DeepAware	RR-Net	DeepAware
Se [%]	97.74	93.06	-	-
Sp [%]	87.10	91.67	95.53	98.47
Acc [%]	88.22	91.82	-	-
FPR [%]	12.90	08.33	04.47	01.53

lenging. Overall, it should be noted that compared to RR-Net, improvements in all three combinations, namely, RR-Net ∧ CAH, RR-Net ∧ DENS-ECG, and DeepAware (RR-Net ∧ DENS-ECG ∧ CAH) are statistically significant.

5. Limitations and future work

The presented *DeepAware* model has three main limitations that require further improvements. First, compared to RRI-based approaches such as the RR-Net model, the proposed *DeepAware* model is computationally expensive. The RR-Net can classify 24 hours of ECG in less than 1 minute, whereas it takes more than 30 minutes for *DeepAware* to analyze the same amount of data in a non-GPU computing environment. Therefore, it may not be straightforward to deploy this model in resource-constrained wearable devices. So, the *DeepAware* model will be more suitable to be used in a cloud computing environment. Secondly, following the limitation of the DENS-ECG model in detecting inverted P-waves, the *DeepAware* model may encounter more FPs in the presence of such ECG morphology with inverted P-waves. To overcome this, the DENS-ECG model should be trained on a dataset with higher number of inverted P-waves morphology, which is currently missing in QTDB dataset [34]. Thirdly, it should also be noted that since the minimum input window length in DeepAware is 30 RRIs, and the minimum P-wave count in a window is kept at 15; therefore the smaller AF segments (e.g., 4–5 seconds) might go undetected. Lastly, although DeepAware helps reducing the FPR and improving the accuracy, its impact on sensitivity (Table 4) under free-living conditions is still a concern that needs further investigation on larger datasets with multi-class arrhythmias. Also, the effectiveness of CAH in the presence of multi-class arrhythmias might change.

In the future work, we plan to extend the CACHET-CADB annotations, evaluate the DeepAware on a diverse dataset with multi-class arrhythmias and work on its interpretability. Besides the heuristics approach, we will also explore employing the ambulatory contexts as direct input features to the DL models.

6. Conclusion

This article presented *DeepAware*, which is a hybrid end-to-end atrial fibrillation detection algorithm that combines deep learning with context-aware heuristics. The model takes three different inputs: (i) RRIs, (ii) raw ECG signals, and (iii) participant's ambulatory context in order to classify AF and non-AF rhythms. Unlike most state-of-the-art models, *DeepAware* has been evaluated on five different datasets, four of which are unseen to the model during the training phase. We found that *DeepAware* is very generalizable and achieve better AF detection performance on public datasets compared to state-of-the-art models. Particularly, the *DeepAware* model performed better when applied on ambulatory ECG collected under free-living conditions. We have also demonstrated that relying only on RRI features for AF detection is problematic, which leads to a high FPR, especially in the presence of confounding arrhythmias (i.e., atrial flutter, PVCs, atrial sinus arrhythmias), and ambulatory motion artifacts from contextual change. The obtained results demonstrate that contextual data col-

lection could be an important factor in improving automatic AF detection in RRI feature-based models under free-living ambulatory conditions. In addition, the *DeepAware* model can significantly reduce the workload required for manual verification of false positives in such longitudinal ambulatory monitoring.

Declaration of Competing Interest

This is to certify that the authors have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakersbureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

CRediT authorship contribution statement

Devender Kumar: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Investigation, Writing – original draft, Writing – review & editing. **Abdolrahman Peimankar:** Conceptualization, Methodology, Software, Writing – review & editing. **Kamal Sharma:** Resources, Writing – review & editing. **Helena Domínguez:** Resources, Writing – review & editing. **Sadasivan Puthusserypady:** Conceptualization, Supervision, Writing – review & editing. **Jakob E. Bardram:** Conceptualization, Data curation, Resources, Supervision, Funding acquisition, Project administration, Writing – review & editing.

Acknowledgments

This work was supported in part by the Innovation fund Denmark under grant # 6153-00009B (REAFEL) and the Copenhagen Center for Health Technology.

References

[1] U.R. Acharya, H. Fujita, S.L. Oh, U. Raghavendra, J.H. Tan, M. Adam, A. Gertych, Y. Hagiwara, Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network, Future Gener Comput Syst 79 (2018) 952–959.
[2] R.S. Andersen, A. Peimankar, S. Puthusserypady, A deep learning approach for real-time detection of atrial fibrillation, Expert Syst Appl 115 (2019) 465–473.
[3] Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE Trans. Neural Networks 5 (2) (1994) 157–166.
[4] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, J Mach Learn Res 13 (2) (2012).
[5] D. Berrar, Cross-validation, in: S. Ranganathan, M. Gribskov, K. Nakai, C. Schönbach (Eds.), Encyclopedia of Bioinformatics and Computational Biology, Academic Press, Oxford, 2019, pp. 542–545.
[6] P.M. Buscema, E. Grossi, G. Massini, M. Breda, F. Della Torre, Computer aided diagnosis for atrial fibrillation based on new artificial adaptive systems, Comput Methods Programs Biomed 191 (2020) 105401.
[7] P.-H. Chan, C.-K. Wong, L. Pun, Y.-F. Wong, M.M.-Y. Wong, D.W.-S. Chu, C.-W. Siu, Head-to-head comparison of the alivecor heart monitor and microlife watchbp office afib for atrial fibrillation screening in a primary care setting, Circulation 135 (1) (2017) 110–112.
[8] X. Chen, Z. Cheng, S. Wang, G. Lu, G. Xv, Q. Liu, X. Zhu, Atrial fibrillation detection based on multi-feature extraction and convolutional neural network for processing ecg signals, Comput Methods Programs Biomed 202 (2021) 106009.
[9] C.C. Cheung, A.D. Krahn, J.G. Andrade, The emerging role of wearable technologies in detection of arrhythmia, Can J Cardiol 34 (8) (2018) 1083–1087.
[10] H. Chougrad, H. Zouaki, O. Alheyane, Deep convolutional neural networks for breast cancer screening, Comput Methods Programs Biomed 157 (2018) 19–30.
[11] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv:1412.3555 (2014).
[12] G.D. Clifford, C. Liu, B. Moody, H.L. Li-wei, I. Silva, Q. Li, A. Johnson, R.G. Mark, Af classification from a short single lead ecg recording: the physionet/computing in cardiology challenge 2017, in: 2017 Computing in Cardiology (CinC), IEEE, 2017, pp. 1–4.
[13] H. Dang, M. Sun, G. Zhang, X. Qi, X. Zhou, Q. Chang, A novel deep arrhythmia-diagnosis network for atrial fibrillation classification using electrocardiogram signals, IEEE Access 7 (2019) 75577–75590.
[14] S.M.P. Dinakarrao, A. Jantsch, M. Shafique, Computer-aided arrhythmia diagnosis with bio-signal processing: a survey of trends and techniques, ACM Comput Surv (CSUR) 52 (2) (2019) 1–37.
[15] Z. Ebrahimi, M. Loni, M. Daneshtalab, A. Gharehbaghi, A review on deep learning methods for ecg arrhythmia classification, Expert Syst Appl (2020) 100033.
[16] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, Y. Li, Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ecg recordings, IEEE J Biomed Health Inform 22 (6) (2018) 1744–1753.
[17] O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, U.R. Acharya, Deep learning for healthcare applications based on physiological signals: a review, Comput Methods Programs Biomed 161 (2018) 1–13.
[18] O. Faust, M. Kareem, A. Ali, E.J. Ciaccio, U.R. Acharya, Automated arrhythmia detection based on rr intervals, Diagnostics 11 (8) (2021) 1446.
[19] O. Faust, A. Shenfield, M. Kareem, T.R. San, H. Fujita, U.R. Acharya, Automated detection of atrial fibrillation using long short-term memory network with rr interval signals, Comput. Biol. Med. 102 (2018) 327–335.
[20] H. Gao, C. Liu, X. Wang, L. Zhao, Q. Shen, E. Ng, J. Li, An open-access ecg database for algorithm evaluation of qrs detection and heart rate estimation, J Med Imaging Health Inform 9 (9) (2019) 1853–1858.
[21] A.S. Go, E.M. Hylek, K.A. Phillips, Y. Chang, L.E. Henault, J.V. Selby, D.E. Singer, Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the anticoagulation and risk factors in atrial fibrillation (atria) study, JAMA 285 (18) (2001) 2370–2375.
[22] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.
[23] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep learning, volume 1, MIT press Cambridge, 2016.
[24] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nat. Med. 25 (1) (2019) 65.
[25] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput 9 (8) (1997) 1735–1780.
[26] A. Jalali, M. Lee, Atrial fibrillation prediction with residual network using sensitivity and orthogonality constraints, IEEE J Biomed Health Inform 24 (2) (2019) 407–413.
[27] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
[28] D. Kumar, J.E. Bardram, Feasibility study of mcardia: A context-aware ecg monitoring system for arrhythmia screening, 2019, Retrieved on Jan 30, 2021 https://www.cachet.dk/research/studies/mcardia,
[29] D. Kumar, R. Maharjan, A. Maxhuni, H. Dominguez, A. Frølich, J.E. Bardram, Mcardia: A Context-Aware ambulatory ECG collectionsystem for arrhythmia screening, *ACM Transactions on Computing for Healthcare (HEALTH)* 3, no. 2 (2022): 1-28. doi:10.1145/3494581.
[30] D. Kumar, S. Puthusserypady, J.E. Bardram, CACHET-CADB, 2021, Retrieved on Sept 28, 2021, doi:10.11583/DTU.14547264.v1.
[31] D. Kumar, S. Puthusserypady, H. Dominguez, K. Sharma, J.E. Bardram, An investigation of the contextual distribution of false positives in a deep learning-based atrial fibrillation detection algorithm, Expert Syst Appl (2021). In press
[32] M. Kumar, R.B. Pachori, U.R. Acharya, Automated diagnosis of atrial fibrillation ecg signals using entropy features extracted from flexible analytic wavelet transform, Biocybern Biomed Eng 38 (3) (2018) 564–573.
[33] Y. Kutlu, D. Kuntalp, A multi-stage automatic arrhythmia recognition and classification system, Comput Biol Med 41 (1) (2011) 37–45, doi:10.1016/j.compbiomed.2010.11.003.
[34] P. Laguna, R.G. Mark, A. Goldberg, G.B. Moody, A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg, in: Computers in cardiology 1997, IEEE, 1997, pp. 673–676.
[35] D. Lai, Y. Bu, Y. Su, X. Zhang, C.-S. Ma, Non-standardized patch-based ecg lead together with deep learning based algorithm for automatic screening of atrial fibrillation, IEEE J Biomed Health Inform 24 (6) (2020) 1569–1578.
[36] D. Lai, X. Zhang, Y. Zhang, M.B.B. Heyat, Convolutional neural network based detection of atrial fibrillation combing rr intervals and f-wave frequency spectrum, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 4897–4900.
[37] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[38] Y. LeCun, Y. Bengio, et al., Convolutional networks for images, speech, and time series, Handbook Brain Theory Neural Netw 3361 (10) (1995) 1995.
[39] M. Limam, F. Precioso, Atrial fibrillation detection and ecg classification based on convolutional recurrent neural network, in: 2017 Computing in Cardiology (CinC), IEEE, 2017, pp. 1–4.
[40] D. Marinucci, A. Sbrollini, I. Marcantoni, M. Morettini, C.A. Swenne, L. Burattini, Artificial neural network for atrial fibrillation identification in portable devices, Sensors 20 (12) (2020) 3570.
[41] M.F. Márquez, L. Colín, M. Guevara, P. Iturralde, A.G. Hermosillo, Common electrocardiographic artifacts mimicking arrhythmias in ambulatory monitoring, Am Heart J 144 (2) (2002) 187–197.
[42] R.J. Martis, U.R. Acharya, H. Adeli, H. Prasad, J.H. Tan, K.C. Chua, C.L. Too, S.W.J. Yeo, L. Tong, Computer aided diagnosis of atrial arrhythmia using dimensionality reduction methods on transform domain representation, Biomed Signal Process Control 13 (2014) 295–305.

[43] S.M. Mathews, C. Kambhamettu, K.E. Barner, A novel application of deep learning for single-lead ecg classification, Comput Biol Med 99 (2018) 53–62, doi:10.1016/j.compbiomed.2018.05.013.
[44] G. Moody, A new method for detecting atrial fibrillation using rr intervals, Comput Cardiol (1983) 227–230, doi:10.13026/C2MW2D.
[45] G.B. Moody, R.G. Mark, The impact of the mit-bih arrhythmia database, IEEE Eng Med Biol Mag 20 (3) (2001) 45–50.
[46] S. Mousavi, F. Afghah, U.R. Acharya, Han-ecg: an interpretable atrial fibrillation detection model using hierarchical attention networks, Comput Biol Med 127 (2020) 104057.
[47] S. Mousavi, F. Afghah, A. Razi, U.R. Acharya, Ecgnet: learning where to attend for detection of atrial fibrillation with deep visual attention, in: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), IEEE, 2019, pp. 1–4.
[48] Movisens, Ecgmove 4 - ecg and activity sensor, 2019, Retrieved on Sept 8, 2021 https://www.movisens.com/en/products/ecg-sensor/.
[49] Movisens, Dataanalyzer – documentation and algorithms, 2021b, Retrieved on Sept 8, 2021 https://docs.movisens.com/DataAnalyzer/#starting-dataanalyzer.
[50] Movisens, Dataanalyzer - sensor data analysis, 2021a, Retrieved on Sept 8, 2021 https://www.movisens.com/en/products/dataanalyzer/.
[51] F. Murat, F. Sadak, O. Yildirim, M. Talo, E. Murat, M. Karabatak, Y. Demir, R.-S. Tan, U.R. Acharya, Review of deep learning-based atrial fibrillation detection studies, Int J Environ Res Public Health 18 (21) (2021) 11302.
[52] Y. Nesterov, A method of solving a convex programming problem with convergence rate o(1/k)2, Soviet Mathematics Doklady 27 (2) (1983) 372–376.
[53] K. O'Shea, R. Nash, An introduction to convolutional neural networks, arXiv preprint arXiv:1511.08458 (2015).
[54] J. Oster, G.D. Clifford, Impact of the presence of noise on rr interval-based atrial fibrillation detection, J Electrocardiol 48 (6) (2015) 947–951.
[55] J. Pan, W.J. Tompkins, A real-time qrs detection algorithm, IEEE Trans Biomed Eng (3) (1985) 230–236.
[56] A. Parsi, M. Glavin, E. Jones, D. Byrne, Prediction of paroxysmal atrial fibrillation using new heart rate variability features, Comput Biol Med 133 (2021) 104367, doi:10.1016/j.compbiomed.2021.104367.
[57] A. Peimankar, S. Puthusserypady, Dens-ecg: a deep learning approach for ecg signal delineation, Expert Syst Appl 165 (2021) 113911, doi:10.1016/j.eswa.2020.113911.
[58] G. Petmezas, K. Haris, L. Stefanopoulos, V. Kilintzis, A. Tzavelis, J.A. Rogers, A.K. Katsaggelos, N. Maglaveras, Automated atrial fibrillation detection using a hybrid cnn-lstm network on imbalanced ecg datasets, Biomed Signal Process Control 63 (2021) 102194.
[59] A. Petrėnas, V. Marozas, L. Sörnmo, Low-complexity detection of atrial fibrillation in continuous long-term monitoring, Comput Biol Med 65 (2015) 184–191, doi:10.1016/j.compbiomed.2015.01.019.
[60] W.H. Press, S.A. Teukolsky, Savitzky-golay smoothing filters, Comput Phys 4 (6) (1990) 669–672.
[61] S. Ramkumar, N. Nerlekar, D. D'Souza, D.J. Pol, J.M. Kalman, T.H. Marwick, Atrial fibrillation detection using single lead portable electrocardiographic monitoring: a systematic review and meta-analysis, BMJ Open 8 (9) (2018) e024178.
[62] D. Raví, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G.-Z. Yang, Deep learning for health informatics, IEEE J Biomed Health Inform 21 (1) (2017) 4–21, doi:10.1109/JBHI.2016.2636665.
[63] T.N. Sainath, O. Vinyals, A. Senior, H. Sak, Convolutional, long short-term memory, fully connected deep neural networks, in: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 4580–4584, doi:10.1109/ICASSP.2015.7178838.
[64] J. Selder, L. Breukel, S. Blok, A. van Rossum, I. Tulevski, C. Allaart, A mobile one-lead ecg device incorporated in a symptom-driven remote arrhythmia monitoring program. the first 5,982 hartwacht ecgs, Netherlands Heart J 27 (1) (2019) 38–45.
[65] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).
[66] G. Tuboly, G. Kozmann, O. Kiss, B. Merkely, Atrial fibrillation detection with and without atrial activity analysis using lead-i mobile ecg technology, Biomed Signal Process Control 66 (2021) 102462.
[67] H.-A. Tzou, S.-F. Lin, P.-S. Chen, Paroxysmal atrial fibrillation prediction based on morphological variant p-wave analysis with wideband ecg and deep learning, Comput Methods Programs Biomed 211 (2021) 106396.
[68] J. Wang, A deep learning approach for atrial fibrillation signals classification based on convolutional and modified elman neural network, Future Gener Comput Syst 102 (2020) 670–679.
[69] X. Wei, J. Li, C. Zhang, M. Liu, P. Xiong, X. Yuan, Y. Li, F. Lin, X. Liu, Atrial fibrillation detection by the combination of recurrence complex network and convolution neural network, J Probab Stat 2019 (2019).
[70] Z. Wu, X. Ding, G. Zhang, X. Xu, X. Wang, Y. Tao, C. Ju, A novel features learning method for ecg arrhythmias using deep belief networks, in: 2016 6th International Conference on Digital Home (ICDH), IEEE, 2016, pp. 192–196.
[71] Y. Xia, N. Wulan, K. Wang, H. Zhang, Detecting atrial fibrillation by deep convolutional neural networks, Comput Biol Med 93 (2018) 84–92.
[72] S.S. Xu, M.-W. Mak, C.-C. Cheung, Towards end-to-end ecg classification with raw signal extraction and deep neural networks, IEEE J Biomed Health Inform 23 (4) (2018) 1574–1584.
[73] X. Xu, S. Wei, C. Ma, K. Luo, L. Zhang, C. Liu, Atrial fibrillation beat identification using the combination of modified frequency slice wavelet transform and convolutional neural networks, J Healthc Eng 2018 (2018).
[74] Z. Yao, Z. Zhu, Y. Chen, Atrial fibrillation detection by multi-scale convolutional neural networks, in: 2017 20th International Conference on Information Fusion (Fusion), IEEE, 2017, pp. 1–6.
[75] Ö. Yildirim, A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification, Comput Biol Med 96 (2018) 189–202.
[76] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: residual learning of deep cnn for image denoising, IEEE Trans Image Process 26 (7) (2017) 3142–3155.
[77] Z. Zhong, L. Jin, Z. Xie, High performance offline handwritten chinese character recognition using googlenet and directional feature maps, in: 2015 13th International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2015, pp. 846–850.