Computer Methods and Programs in Biomedicine 000 (2021) 105931

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Adaptive learning and cross training improves R-wave detection in ECG

Nagarajan Ganapathy [a,b], Ramakrishnan Swaminathan [b], Thomas M. Deserno [a,*]

[a] Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, 38106 Braunschweig, Germany
[b] Biomedical Engineering Group, Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, 600036 Tamil Nadu, India

ARTICLE INFO

Article history:
Received 13 November 2019
Accepted 4 January 2021
Available online xxx

Keywords:
Electrocardiography
R-wave detection
Deep learning
Convolutional neural network
Cross-database training
Cross-lead training
Adaptive training
Adaptive model

ABSTRACT

Background and Objective: Automated R-wave detection plays a vital role in electrocardiography (ECG) and ECG-based computer-aided diagnosis. Recently, a multi-level one-dimensional (1D) deep learning approach was presented that shows good performance as compared to traditional methods.

Methods: In this paper, we present several improvements of the multi-level 1D convolutional neural network (CNN)-based deep learning approach using: (i) adaptive deep learning, (ii) cross-database training, and (iii) cross-lead training. For this, we consider ECG signals from four publicly available databases: MIT-BIH, INCART, TELE, and SDDB, having 109,404, 175,660, 6,708, and 1,684,447 annotated beats, respectively. Except for TELE, all databases provide at least two-lead recordings. To evaluate the improvements, experiments are performed with adaptive k-times cross-trained databases validation scheme (k = 5). The hypothesis tested are: (i) the improvements outperform the state-of-the-art, (ii) cross-database training and adaptive deep learning contribute, and (iii) additional databases or cross-lead training further improves the results.

Results: Our proposed approach outperforms the state-of-the-art. In terms of F-measure, F = 99.75% and F = 95.25% is obtained for the MIT-BIH and TELE databases, respectively. Further, cross-database training (F = 98.02%) is found to be more effective than training on individual databases (F = 97.33%). The performance of our approach further improves when additional databases and different leads are used for training.

Conclusion: Existing state-of-the-art methods perform low on noisy and pathological signals. Adaptive cross-data training identifies the optimal model. Using multiple datasets and leads allows analyzing noisy, pathological and mobile-recorded long-term ECG signals without ground truths. These conclusions are based on the comprehensive evaluation of four different databases, and in total, about 4.5 million annotated beats.

1. Introduction

Cardiovascular diseases (CVDs) are considered as the leading cause of death. According to the World Health Organization, 17.7 million people died from CVDs in 2015, and 37% were premature deaths. It is further expected that CVDs will remain in the first place up to 2035 [1,2].

Electrocardiography (ECG) represents the bioelectrical activity of the heart or cardiac muscle in an individual heartbeat. An ECG is the basis of a simple, noninvasive, and well-established CVD diagnosis [3]. Anomalies in morphological patterns of the ECG waveform indicate heart diseases. However, the characteristics and time-varying dynamics of ECG patterns are highly complex and significantly different, even for a normal subject. The morphological characteristics also vary under different physiological and pathological conditions. Furthermore, artifacts, noise, dropouts, and class imbalances increase the complexity of ECG analysis [4].

Manual analyses use a plot of just a few successive cardiac cycles. Clinically, the ECG is analyzed over a short period using R-wave detection and beat decomposition methods. For long-term ECG, Holter monitoring is performed with multiple leads over 12 to 24 hours. With the advancement of wearable sensors, ECG signals are nowadays recorded for more than 24 hours [5], and manual examination is tedious and time-consuming.

* Corresponding author.
E-mail addresses: nagarajan.ganapathy@plri.de (N. Ganapathy), sramki@iitm.ac.in (R. Swaminathan), thomas.deserno@plri.de (T.M. Deserno).

https://doi.org/10.1016/j.cmpb.2021.105931
0169-2607/

Therefore, automated and robust R-wave detection is necessary and remains an intense area of research for decades [5]. Further, automated R-wave detection and heartbeat classification is also required to classify arrhythmia and heart abnormalities [6]. Various method including filter banks [7], derivatives [8,9], transforms [10–12], statistical and morphological approaches [13,14], threshold differences [15,16], neural network [17,18], and mixture of hybrid experts [19] have been proposed. Most of the methods were evaluated with a specific database and, thus, have limited generalizability. In addition, these methods have a high variation in performance when applied to signals acquired in different conditions, or from multi-morbid patients. In clinical practice, the accuracy reduces further due to inter- and intra-patient variations [20,21].

In order to adapt generalizability, Kashif, Jonas and Deserno (Kashif) have proposed a simultaneous truth and performance level estimation (STAPLE) approach to combine nine R-wave detectors [2]. The STAPLE combined method performs better than the best individual algorithm. The authors evaluated their approach using public and private databases [2]. The reported performance was F = 99.73% and F = 97.60% for MIT-BIH and TELE databases, respectively.

Recently, several deep learning (DL)-based patient-specific ECG classification schemes have been used for heart disease detection [22,23]. For instance, R-wave localization is used for cost-effective and accurate population-specific screenings in real-time scenarios [24]. DL-based methods automatically learn and extract critical features from the ECG and provide an abstract representation of the signal [25]. In a study, Chandra et al. (Chandra) have used single layer CNN network to detect R-Peak locations [26]. Moreover, The DL methods have also been used for the classification of pathological conditions such as atrial fibrillation, ventricular fibrillation and congestive heart disease [22,26]. However, most DL approaches do not identify the particular waves within the recording.

In 2018, Xiang, Lin, and Meng (Xiang) proposed a DL-based method for accurate QRS complex detection [3]. The network contains a hierarchical parallel two-level one-dimensional (1D) convolutional neural network (CNN), each with varying network parameters. In level one, a two-layer 1D-CNN is employed to extract robust features from ECG segments, while in the second level, a single layer 1D-CNN obtains the abstract features. The extracted features are concatenated and applied to a fully connected layer to distinguish QRS and non-QRS segments. For training, a small number of QRS and non-QRS segments are used. The evaluation on MIT-BIH [27] and INCART [28] databases, both resampled to 360 Hz, yields sensitivity and specificity of 99.77% and 99.86%, respectively; a slight improvement of the STAPLE approach. However, the performance of an MIT-BIH trained network applied to the INCART data or vice versa is not assessed. This, however, would be a far more realistic scenario, since a trained network is applied to data recorded by different mobile devices.

To improve the Xiang approach, we propose (i) adaptive DL to support patient-specific models, and (ii) cross-database training as well as (iii) cross-lead training to increase the number and variety of signals seen in the learning period.

2. Methods

In this section, we briefly recapitulate the approach of Xiang and then describe our improvements.

2.1. The approach of Xiang

Xiang uses a two-level 1D-CNN method. The ECG signals are pre-processed using difference signals (subtraction between adjacent samples (Diff)) and average difference signals (averaged per several adjacent difference samples (avgDiff)) to characterize the high slope of QRS complexes. The local-level features are obtained from the difference signal to provide more detailed and localized descriptors for the signals. Accordingly, global features are extracted from an average difference of the input signal (Fig. 1). The set of parameters, such as the number of layers, learning rate and epochs are considered from the Xiang approach. On the other hand, the network parameters such as kernel size, filters, and momentum are determined to obtain an optimal performance with minimum training error. Using random search technique, the CNN parameters namely filters is set as 32, 16, kernel is set to 5, dropout is set to 0.5, and momentum is initialized as 0.1, respectively.

For training, snippets of fixed length representing QRS and non-QRS segments are taken, based on the annotations of the reference databases [29]. In accordance with the EC57 and ANSI/AAMI EC38 standards, each snippet consists of 56 sampling points: 23 points before and 33 points after the R-wave position [30]. To maintain uniformity across the database, ECG signals are filtered using a 512-order lowpass finite impulse response filter with a cutoff frequency of 180 Hz and later down-sampled to 360 Hz (Fig. 2). The finite impulse response filter with cut-off frequency computed using Nyquist theorem is employed to avoid aliasing [31,32].

2.2. Improved framework

Our method consists of four major steps: (i) ECG preprocessing as of Xiang, (ii) cross-data and cross-lead training, (iii) adaptive QRS complex prediction, and (iv) R-wave detection as of Xiang. (Fig. 3).

2.3. Cross-database training

The performance of the network depends on the training data and DL approaches require large sets of such data, which further must carefully be separated from the testing data. Furthermore, the selection of training data plays a vital role in computer-aided diagnosis and decision making [33]. The trained model performs low if the quantity or quality of data in training is inappropriate. The performance is also low on data that is not represented during training. We observed that increased variability of input during the learning phase does not necessarily yield a drop-in performance. Instead, combining a variety of training sets from multiple databases improves the generalizability of the model.

2.4. Deep learning based adaptive QRS detection

Identifying superior features or training sets increases the sensitivity of a computer-aided diagnostic system or vice-versa [4,34]. Identification and selection of better training data is a complex task and depends on various factors such as type, size, origin, and quality of data. Furthermore, complexity is increased if multivariate clinical databases are used for analysis. In most of the published approaches, cross-validation techniques such as k-fold and leave-one-out techniques have been used to determine the training data [3,4,35,36]. However, adaptive training for multivariate clinical data has not yet been proposed.

We suggest a STAPLE-based adaptive method to identify the optimal training model automatically from multiple training sets. STAPLE is popular in medical image segmentation and already applied to reliable R-wave detection [2]. It is a vote based ranking based method which identify the trained model that has better performance for all the validation sets. Kashif have used STAPLE to determine R-wave by voting across nine R-wave detection algorithms [2]. Motivated by this approach, in our study, we used STAPLE method to determine the robust training model which performs equally better across all the training sets. This is referring to

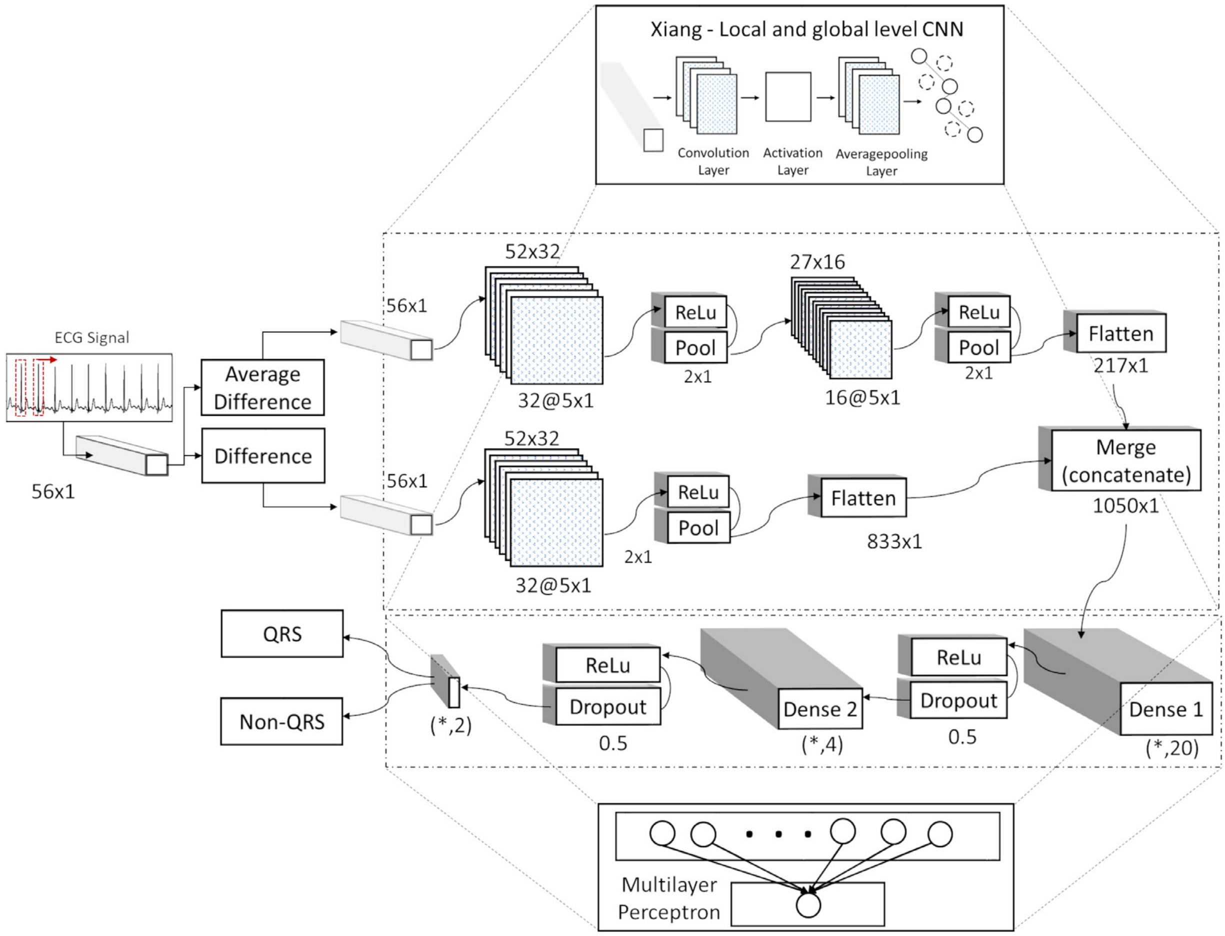

Fig. 1. Xiang proposed a two-level 1D-CNN model that is composed of two parallel layers, each comprising 32 filters with a kernel size of 5 × 1.

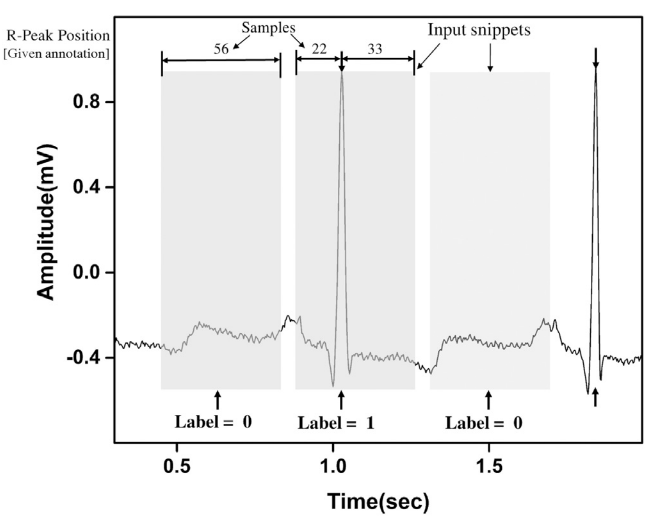

Fig. 2. The QRS snippets are labeled from the annotated ground truth.

as selection of adaptive training models. These identified adaptive trained deep learning model is finally employed to robustly locate R-waves in test sets. These adaptive trained models can be further improvised with cross-data training to include multiple datatypes, channels and quality of the signals.

Thus, we generate multiple patient-specific training models for individual databases. Later, the top *n* best models of individual databases identified using mean validation accuracy are combined to generate an effective cross-trained model, without any repetition. The combined models then identify the QRS complexes. To improve the reliability and robustness of the Xiang approach, the STAPLE algorithm is further used to determine the most common QRS location from an ensemble of k cross-trained adaptive models. For this study, random training samples are obtained from the individual dataset using k-fold cross validation; here, k is set to 5 for cross data-trained adaptive training model. In this cross-validation technique, the samples are divided into k uniformly random sets, the k-1 set is used for training and the remaining set is used as a testing sample.

2.5. Cross-lead training

Different ECG datasets can be collected using multiple devices with varying leads and different experimental protocols. These technical variations will add discrepancies to the existing dataset across the subjects. Hence, to reduce inter- and intra-subject variability across the database, cross-data training and adaptive model selection are further enhanced by including cross-lead training [37]. In this study, multiple ECG leads are considered as the part of input data for training sets only, if such leads exhibit similar morphological shapes in ECG waveforms, e.g., I, II, V5, V6, AVL. These additional samples are randomly selected to explore the benefits of cross-lead variability.

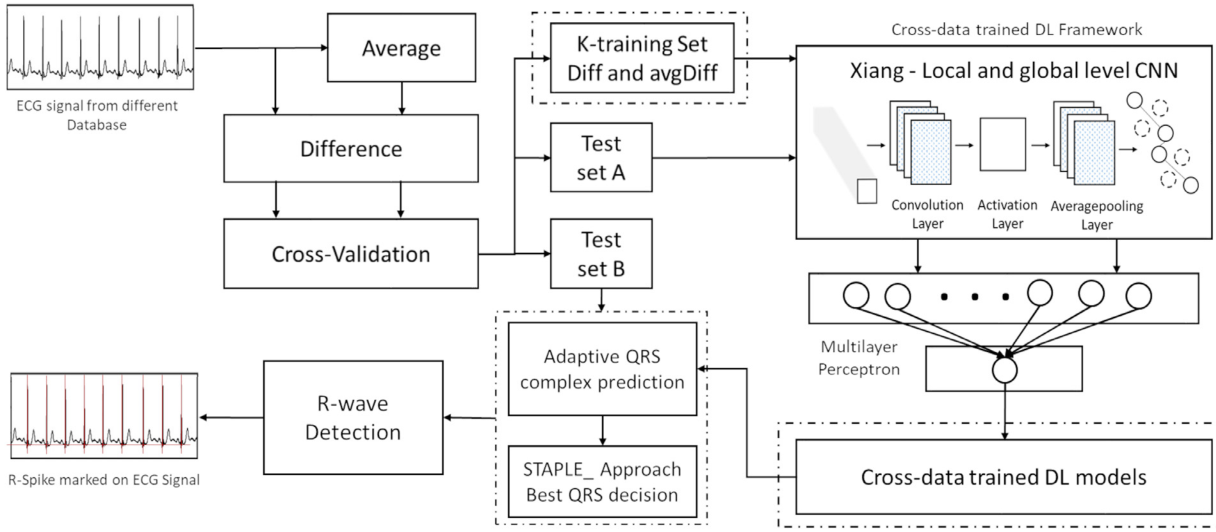

Fig. 3. Block diagram of our method. The dotted lines indicate the extensions to the approach of Xiang [3] .

3. Evaluation

As state-of-the-art, we consider the STAPLE-based approach of Kashif [2] as well as the non-adaptive and single database-trained DL model of Xiang [3]. The STAPLE method employs nine individual R-wave detectors namely, Arzeno et al. [8], Chernenko [38], Arteaga-Falconi et al. [9], Liu et al. [10], Manikandan et al. [39], Pan and Tompkins [40], Khamis et al. [41], Afonso et al. [7], and Madeiro et al. [42].

3.1. Hypothesis

H1: Our method outperforms state-of-the-art R-wave detection.

H2: The performance gain is due to both cross-database and adaptive training.

H3: Additional training data or cross-lead training or further improves performance.

3.2. Experiments

In order to prove (or disapprove) the hypothesis, three experiments are designed and conducted:

E1: The performance of Kashif, Chandra, Xiang, and our method (ExtX) are evaluated using manual ground truth. In addition, we extend Kashif approach as follows: (i) Kashif + Xiang, (ii) Kashif + ExtX, and (iii) Kashif + Xiang + ExtX. Hypothesis H1 will be accepted if ExtX or Kashif extensions along with ExtX perform best. In order to compare with previous papers, we compute the results for all databases individually and consider lead II only.

E2: We extend Xiang with (i) the adaptive component, (ii) the cross-data training, and (iii) both, which equals ExtX. The hypothesis H2 is accepted if the performance gains for (i) to (iii).

E3: We extend the data that is used for training by feeding several leads individually. Also, we use additional data annotated with the appropriate ground truth. H3 is accepted if these extensions also yield a performance gain.

3.3. Databases

Although there are several ECG databases that have been used by researchers towards R-wave detection, most use MIT-BIH Arrhythmia [27] and St. Petersburg Institute of Cardiological Technics 12-lead Arrhythmia (INCART) [28]. Further, INCART and TELE [42] contain multi-morbid and remotely monitored signals with most common heart abnormalities.

Therefore, we have used MIT-BIH, INCART and TELE database to test and validate the robustness of our proposed method (experiment E1). For experiment E2, we merge these databases. To perform experiment E3, additional data is required. We use the Sudden Cardiac Death Holter Database (SDDB) as it contains multimorbid long-term ECG signals [24,43]. To compare the results, SDDB is used only for additional adaptive training. In SDDB, each record provides a different set of leads. Three records (out of the 23 available) do not have any of the useful leads L1=I, L2=II, V5, V6, or AVL and hence, they have been excluded. Furthermore, we use L1=I and L2=II of MIT-BIH as well as L1=I, L2=II, V5, V6, AVL of INCART.

Altogether, the total number of annotated beats sums up to approximately 2 million (Table 1). In the experiment E3, the total number of beats is 4.5 million. Extended training was done in three parts: Multiple lead training was applied, SDDB has been added for training with a single lead, and SSDB data was added for training with multiple leads.

3.4. Metrics

Qualitative and quantitative evaluation is used to analyze the performance of our method. Based on the previous publication,

Table 1
Databases in use for the experiments. All data is publicity freely available, annotated with ground truth of R-wave localization, and contains pathological records.

Database	Leads	Subject	Records	Sampling rate [Hz]	Duration [min]	Total length [beats]
MIT-BIH	2	47	48	360	30 ± 0	109,404
INCART	12	32	75	257	30 ± 0	175,000
TELE	1	120	250	500	0.48 ± 0.24	6,708
SDDB	2	20	20	250	1115 ± 468	1,684,447

Table 2
Overall performance [%] (mean and standard deviation (Std.)) of R-wave detection methods with training and evaluation on same database.

Database	MIT-BIH						INCART						TELE					
	Precision		Recall		F-Measure		Precision		Recall		F-Measure		Precision		Recall		F-Measure	
Method	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Kashif	99.93	0.24	99.06	2.39	99.48	1.34	99.82	0.63	98.24	3.89	98.99	2.27	82.80	27.22	89.23	25.17	84.76	26.23
Chandra	99.96	0.08	99.36	1.60	99.65	0.82	98.05	2.68	86.58	27.28	88.52	25.13	95.77	7.57	**94.92**	12.56	94.91	9.65
Xiang	99.63	0.70	99.69	0.89	99.65	0.65	98.17	2.50	99.17	2.85	98.62	2.11	97.17	6.60	93.63	14.15	94.64	10.55
ExtX	99.62	0.99	**99.88**	0.38	**99.75**	0.55	98.53	2.58	99.54	1.96	99.00	1.72	**98.09**	8.48	93.72	15.16	**95.25**	12.04
Kashif + Xiang	99.96	0.13	99.36	1.60	99.64	0.87	99.87	0.51	98.69	2.89	99.25	1.67	85.55	24.87	91.69	22.08	87.56	23.59
Kashif + ExtX	99.96	0.13	99.38	1.54	99.66	0.83	99.88	0.50	98.74	2.83	99.28	1.64	85.60	24.83	91.81	22.00	87.65	23.54
Kashif + Xiang+ ExtX	**99.98**	0.06	99.43	1.45	99.69	0.76	**99.94**	0.25	**98.89**	2.63	**99.39**	1.46	88.34	23.79	92.51	21.62	89.54	22.82

measures namely, precision (P), recall (R), F-measure (F), and score (S) are computed [36].

4. Results

E1: The best F-measures of 99.75% and 95.25% for MIT-BIH and TELE, respectively, are obtained using the ExtX approach (Table 2). For INCART data, the STAPLE method combined with Xiang and ExtX yields the best performance (F = 99.39%). For TELE data, the top three methods: ExtX, Chandra and Xiang obtained the F-measures 99.25%, 94.91% and 94.64%, respectively. The ExtX approach yields the highest recall of 99.88% for MIT-BIH. Therefore, hypothesis H1 is confirmed.

E2: In all experiments, the adaptive training outperforms the non-adaptive one (Table 3). The best performance is obtained for the combined database. Here, the average F-measures of adaptive and non-adaptive training yields 98.02% and 97.33%, respectively. The adaptive training method achieves the highest recall of 98.16% for TELE. Except, MIT-BIH, the precision of the adaptive approach is higher than 95.00%. This confirms our hypothesis H2.

E3: Multiple lead training improves single lead training with or without additional databases and adding annotated ground truth data in training increases performance for a single as well as multiple lead training (Table 4). The performance of the extended training is F-measure: 99.08% for multiple lead training and additional databases. The recall rate of the our method with multiple leads is found to be high (99.49% and 99.43%) than single lead for combined databases. This confirms our hypothesis H3.

Except for TELE, the training accuracy per epochs is above 90% for both individual and cross-database (see Fig. 4a). Similarly, the training loss is observed to be less than 20% for early epochs and reduces with increased epochs (see Fig. 4b). The training loss of combined database is low at early epochs and remains constant.

5. Discussion

For decades, automated R-wave detection is an important topic of research. Although several algorithms have been proposed and authors have claimed accuracies of more than 99%, novel approaches are being developed to address the issues associated with computational electrocardiography [5,6]. Specifically, generalizability and robustness of existing approaches with several ECG databases, mobile data recording, and low-quality signals from inexpensive wellness devices are not tested comprehensively, yet [5].

In recent years, several machine learning approaches have been proposed to diagnose heart diseases; however machine learning is not extensively used for R-wave detection [24]. The reported performances of these methods depend on the selection of samples used for training and validation. Similarly, the coarse-fine grained features from ECG signals can be obtained from the sub-bands using wavelet transform. However, there are reports that describe the selection of mother wavelet in wavelet transform plays a vital role for better features. Also, the analysis of signals using wavelet transform is computationally expensive [12]. In addition, advancements in wearable sensor technology may provide lifetime ECG monitoring, but requires a robust and patient-specific approach to handle dynamic characteristics of long-term ECG data [2–4,36].

In this paper, a versatile cross-data trained adaptive DL approach for robust R-wave detection is presented, which is suitable for multiple databases and multi-parametric ECG signals. To the best of our knowledge, the idea of cross-data training and adaptive model selection for DL is first of its kind. The experimental evaluations show that our methods can be a reliable means for robust R-wave detection. With the support of adaptive training, our methods can be trained for an individual patient, resulting the R-wave detection to be more patient-specific.

Table 3
Average performance metrics [%] (mean and standard deviation) of non-adaptive and adaptive training.

	Non-adaptive						Adaptive					
	Precision		Recall		F-Measure		Precision		Recall		F-Measure	
Database	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
MIT-BIH	90.96	22.68	70.15	33.45	74.80	30.60	91.34	13.86	88.05	16.64	**87.68**	15.53
INCART	92.71	14.23	90.17	20.42	90.35	17.96	96.07	8.84	91.43	12.91	**92.15**	10.97
TELE	96.84	8.08	94.39	13.84	94.96	10.90	97.20	6.32	98.16	5.65	**97.30**	6.09
MIT-BIH + INCART + TELE	97.96	4.61	97.27	6.13	97.33	5.11	98.77	3.98	97.73	5.79	**98.02**	4.74

Table 4
Average performance metrics [%] of our method using combined databases.

Learning set	Number of leads					
	Single			Multiple		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure
MIT-BIH + INCART + TELE	98.77	97.73	98.02	98.53	99.49	**99.01**
MIT-BIH + INCART + TELE + SDDB	97.96	99.35	98.65	98.73	99.43	**99.08**

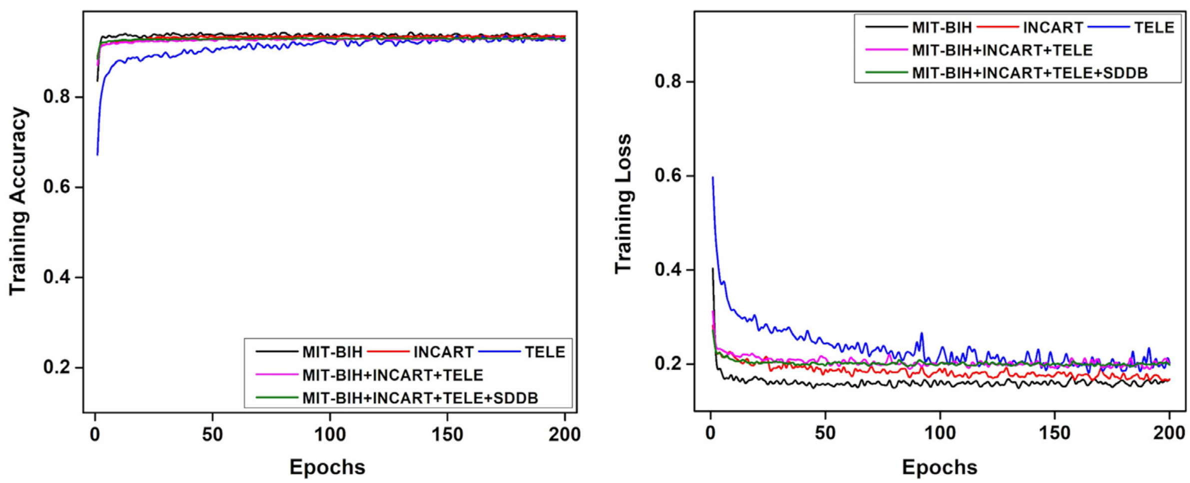

Fig. 4. Mean (a) training accuracy per epochs, and (b) training loss per epochs for MIT-BIH, INCART, TELE database and its cross-datasets.

Therefore, with our extensions, the R-wave detection becomes patient-specific and performs well on multi-variable QRS amplitudes, small QRS complexes, and noisy pathological signals (Fig. 5). Furthermore, it is robust to various pathological signal (Fig. 5 a-e). The figure illustrates the similarity of computed R-wave position and manual ground truth. It is seen that in record 210 of MIT-BIH and I65 of INCART, our method detects abnormal R-wave, which is completely missed by other R-wave detectors (Fig. 5a and 5b). Similarly, in record 203 of MIT-BIH, the extended STAPLE approach and our method determines the abnormal R-wave, which is missed by conventional R-wave method (Fig. 5d). It may be attributed to the robustness of the adaptive DL approach to differentiate normal and abnormal beats. Also, cross-database training improves the models for being more robust to noisy dataset. Based on Fig. 5, it can be observed that our method has consistently given the best performance for considered all database. It is due to the ability of our proposed method along with STAPLE approach to combines the strength of multiple algorithms with DL, and thus improves the R-wave detection ability.

In order to compare the generalizability of our approach with previous reports, the overall performance score S is computed for all the databases. ExtX is robust among several databases with the best overall score: S = 99.72% and S = 95.31% for the MIT-BIH and the TELE databases, respectively (Table 5). Since their scores are lower than 90%, all non-DL methods are ineffective for the TELE database. This may result from noisy and complex characteristics of the signals in the TELE database. Due to the patient-specific capability of our approach, its score is best for TELE. In addition, it is worth mention that on INCART data, Kashif outperforms Xiang.

Table 5
Overall score [%] of R-wave detection methods on multiple databases. ExtX uses all data and leads for training, while Kashif and Xiang are trained only with data of the same database as used in the validation.

Database	MIT-BIH	INCART	TELE
Kashif	99.51	99.01	87.51
Xiang	99.65	98.59	94.62
ExtX	**99.72**	98.96	**95.31**
Kashif + Xiang	99.66	99.26	89.73
Kashif + ExtX	99.67	99.29	89.82
Kashif + Xiang + ExtX	99.71	**99.40**	91.50

Generally, the performance of the R-wave detection algorithm is tuned to the specific database; however, our method is designed to be reliable for multiple databases and various recorders. This is mainly due to its cross-data training and the adaptive learning ability. In line with Zhai and Tin [33], it also provides an automated method for the selection of training samples for better performance. We observed that the performance of ExtX enhances, when more data is used for training, due to eliminating the weakness of individual databases.

Comparing the performance of adaptive and non-adaptive training, adaptive training is found to be superior. This is in accordance with the findings of Kiranyaz et al. [4], who concluded that adap-

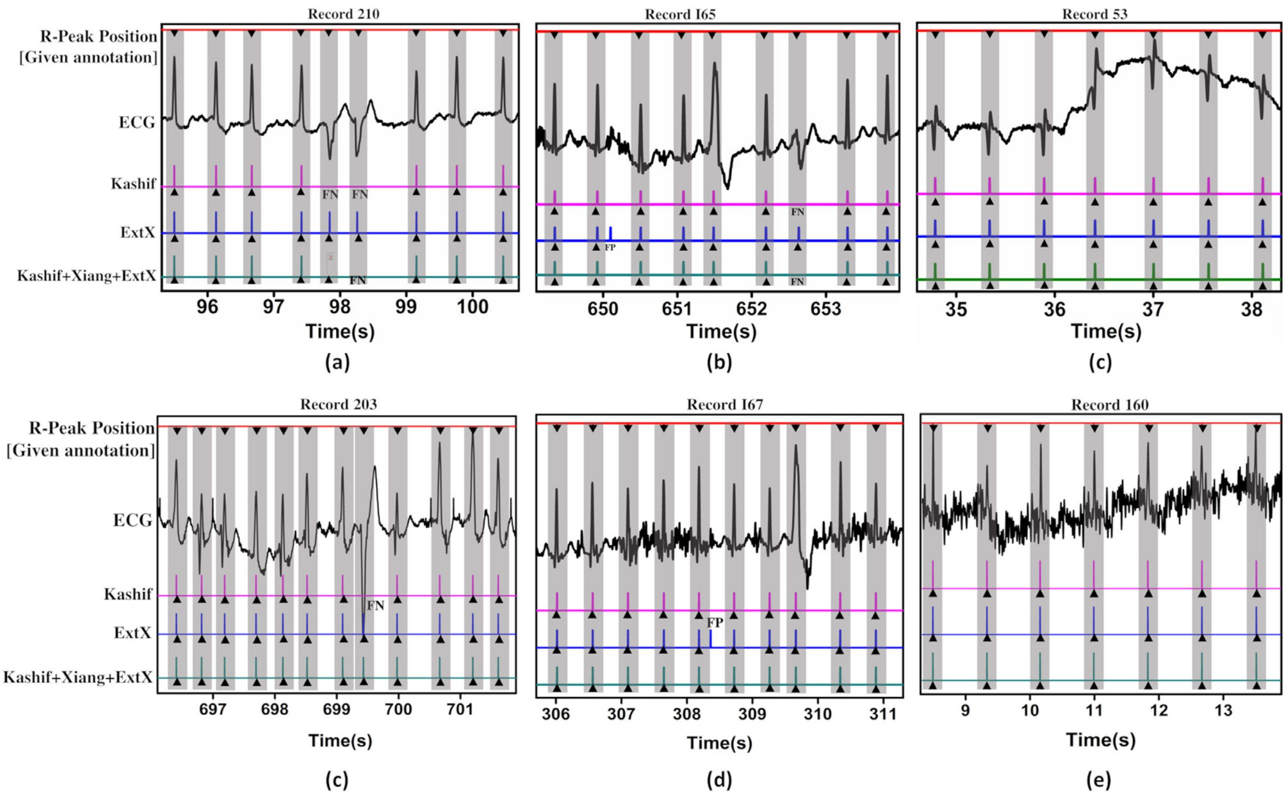

Fig. 5. Robust R-wave detection using the ExtX approach: (a) MIT-BIH Record 210, (b) INCART Record I65, (c) TELE Record 53, (d) MIT-BIH Record 203, (e) INCART Record I67, and (e) TELE Record 160.

tive learning is better for continuous monitoring. This again indicates that with the proposed extensions, 1D-CNN is suitable for regular monitoring, wearable systems and long-term ECG recordings. As the original approach of Xiang is based on a parallel CNN, we assume further that Xiang et al. as well as our method outperform approaches based on a single layer CNN only, such as the work of Chandra et al. [26]. Also, in comparison to single- and parallel-layer CNN, our method is robust across multiple database. Furthermore, our approach can be used for multiple leads.

Our method is found to perform best on MIT-BIH and TELE databases. The work of Xiang has established a milestone in the detection of R-waves in ECG recordings. Extending Xiang's method by (i) cross-database training, (ii) cross-lead training, and (iii) adaptive deep learning, performance increases by 0.08%, 0.4% and 0.7% points for MIT-BIH, INCART, and TELE databases, respectively. Further, the performance of Xiang extended CNN is F = 98.02% and 98.29% for combined database and multiple lead datasets. By counting the number of beats that are used in the different experiments (E2 and E3), it is observed that the massive increase in data contributes only little improvement. Our method overcomes the challenge of training for existing algorithms which identify the best training sets by trial and error. This is why the performance of non-DL methods are found to be highly variable. Indeed, random training samples comprising noisy and multi-morbid patient ECG signals clearly show high variability in performance.

6. Conclusions

In this paper, we extended the Xiang approach using a robust cross-data trained adaptive deep learning approach for R-wave detection that systematically exploits the temporal dependencies and dynamic characteristics of ECG signals. We use a two-level 1D-CNN for adaptive and patient-specific feature generation. In order to maintain robustness across several databases, cross-data and adaptive training are proposed. Our approach yields best F-measure of 99.75%, 95.25% for MIT-BIH and TELE database, respectively. Especially, the best overall score of 99.72% is obtained by MIT-BIH for multiple database training. Cross-data and adaptive training approaches improve performance as compared to conventional methods. In the future, we will face the embedding to real-time monitoring systems and verify the performance in real-world scenarios.

Ethics approval and consent to participate

Not applicable.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions

NG, RS and TD carried out the research. NG developed the method and performed the experiment. NG and TD drafted the manuscript. All authors read and approved the final manuscript.

Data availability

The ECG signals are obtained from the MIT-BIH, INCART, TELE, and SDDB database. All the database are public domain database and available online: https://www.physionet.org/physiobank/database/

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank Mahsa Raeiati, Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Braunschweig, Germany, for her support in MATLAB programming of (34) and Muhammad Kashif, Department of Medical Informatics, RWTH Aachen University, Aachen, Germany for his support in programming the STAPLE algorithm (2). The authors would also like to thank the German Academic Exchange Service (DAAD) for supporting our collaborative research.

References

[1] M. Llamedo, J.P. Martinez, An automatic patient-adapted ECG heartbeat classifier allowing expert assistance, IEEE Trans. Biomed. Eng. 59 (2012) 2312–2320, doi:10.1109/TBME.2012.2202662.
[2] M. Kashif, S.M. Jonas, T.M. Deserno, Deterioration of R-wave detection in pathology and noise: a comprehensive analysis using simultaneous truth and performance level estimation, IEEE Trans. Biomed. Eng. 64 (2017) 2163–2175, doi:10.1109/TBME.2016.2633277.
[3] Y. Xiang, Z. Lin, J. Meng, Automatic QRS complex detection using two-level convolutional neural network, Biomed. Eng. OnLine. 17 (2018) 13, doi:10.1186/s12938-018-0441-4.
[4] S. Kiranyaz, T. Ince, M. Gabbouj, Real-time patient-specific ECG classification by 1-D convolutional neural networks, IEEE Trans. Biomed. Eng. 63 (2016) 664–675, doi:10.1109/TBME.2015.2468589.
[5] T.M. Deserno, N. Marx, Computational electrocardiography: revisiting holter ECG monitoring, Methods Inf. Med. 55 (2016) 305–311, doi:10.3414/ME15-05-0009.
[6] E.J. da S. Luz, W.R. Schwartz, G. Cámara-Chávez, D. Menotti, ECG-based heartbeat classification for arrhythmia detection: a survey, Comput. Methods Programs Biomed. 127 (2016) 144–164, doi:10.1016/j.cmpb.2015.12.008.
[7] V.X. Afonso, W.J. Tompkins, T.Q. Nguyen, S. Luo, ECG beat detection using filter banks, IEEE Trans. Biomed. Eng. 46 (1999) 192–202, doi:10.1109/10.740882.
[8] N.M. Arzeno, Z. Deng, C. Poon, Analysis of first-derivative based QRS detection algorithms, IEEE Trans. Biomed. Eng. 55 (2008) 478–484, doi:10.1109/TBME.2007.912658.
[9] J. Arteaga-Falconi, H.A. Osman, A.E. Saddik, R-peak detection algorithm based on differentiation, in: Proc IEEE WISP, 2015, pp. 1–4, doi:10.1109/WISP.2015.7139157.
[10] X. Liu, J. Yang, X. Zhu, S. Zhou, H. Wang, H. Zhang, A novel R-peak detection method combining energy and wavelet transform in electrocardiogram signal, Biomed. Eng. Appl. Basis Commun. 26 (2014) 1450007, doi:10.4015/S1016237214500070.
[11] A. Sharma, S. Patidar, A. Upadhyay, U.R. Acharya, Accurate tunable-Q wavelet transform based method for QRS complex detection, Comput. Electr. Eng. 75 (2019) 101–111, doi:10.1016/j.compeleceng.2019.01.025.
[12] P.S. Addison, Wavelet transforms and the ECG: a review, Physiol. Meas. 26 (2005) R155–R199, doi:10.1088/0967-3334/26/5/R01.
[13] F. Zhang, Y. Lian, QRS detection based on multiscale mathematical morphology for wearable ECG devices in body area networks, IEEE Trans. Biomed. Circuits Syst. 3 (2009) 220–228, doi:10.1109/TBCAS.2009.2020093.
[14] S.-W. Chen, H.-C. Chen, H.-L. Chan, A real-time QRS detection method based on moving-averaging incorporating with wavelet denoising, Comput. Methods Programs Biomed. 82 (2006) 187–195, doi:10.1016/j.cmpb.2005.11.012.
[15] Y.-C. Yeh, W.-J. Wang, QRS complexes detection for ECG signal: the difference operation method, Comput. Methods Programs Biomed. 91 (2008) 245–254, doi:10.1016/j.cmpb.2008.04.006.
[16] I.I. Christov, Real time electrocardiogram QRS detection using combined adaptive threshold, Biomed. Eng. OnLine. (2004) 9.
[17] B. Abibullaev, H.D. Seo, A new QRS detection method using wavelets and artificial neural networks, J. Med. Syst. 35 (2011) 683–691, doi:10.1007/s10916-009-9405-3.
[18] M.B.I. Reaz, L.S. Wei, Detection of the R wave peak of QRS complex using neural network, in: Proc ICTTA, IEEE, 2004, p. 381, doi:10.1109/ICTTA.2004.1307790.
[19] C. Meyer, J.F. Gavela, M. Harris, Combining algorithms in automatic detection of QRS complexes in ECG signals, IEEE Trans. Inf. Technol. Biomed. 10 (2006) 468–475, doi:10.1109/TITB.2006.875662.
[20] P. de Chazal, M. O'Dwyer, R.B. Reilly, Automatic classification of heartbeats using ECG morphology and heartbeat interval features, IEEE Trans. Biomed. Eng. 51 (2004) 1196–1206, doi:10.1109/TBME.2004.827359.
[21] U. Satija, B. Ramkumar, M.S. Manikandan, A review of signal processing techniques for electrocardiogram signal quality assessment, IEEE Rev. Biomed. Eng. 11 (2018) 36–52, doi:10.1109/RBME.2018.2810957.
[22] P. Bizopoulos, D. Koutsouris, Deep learning in cardiology, IEEE Rev. Biomed. Eng. 12 (2019) 168–193, doi:10.1109/RBME.2018.2885714.
[23] A. Rizwan, A. Zoha, I. Mabrouk, H. Sabbour, A. Al-Sumaiti, A. Alomaniy, M. Imran, Q. Abbasi, A review on the state of the art in atrial fibrillation detection enabled by machine learning, IEEE Rev. Biomed. Eng. (2020) 1–1, doi:10.1109/RBME.2020.2976507.
[24] H. Abrishami, M. Campbell, C. Han, R. Czosek, X. Zhou, P-QRS-T localization in ECG using deep learning, in: 2018 IEEE EMBS Int. Conf. Biomed. Health Inform, BHI, 2018, pp. 210–213, doi:10.1109/BHI.2018.8333406.
[25] L. Deng, D. Yu, Deep learning: methods and applications, Found. Trends Signal Process. 7 (2014) 197–387, doi:10.1561/2000000039.
[26] B.S. Chandra, C.S. Sastry, S. Jana, S. Patidar, Atrial fibrillation detection using convolutional neural networks, 2017 Comput. Cardiol. CinC (2017) 1–4, doi:10.22489/CinC.2017.163-226.
[27] G.B. Moody, R.G. Mark, The impact of the MIT-BIH arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (2001) 45–50, doi:10.1109/51.932724.
[28] 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 (2000) E215–E220.
[29] C. Ye, B.V.K.V. Kumar, M.T. Coimbra, Heartbeat classification using morphological and dynamic features of ECG signals, IEEE Trans. Biomed. Eng. 59 (2012) 2930–2941, doi:10.1109/TBME.2012.2213253.
[30] ECAR AAMIRecommended Practice for Testing and Reporting Performance Results of Ventricular Arrhythmia Detection Algorithms, Association for the Advancement of Medical Instrumentation, 1986.
[31] 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 (2018) 1744–1753, doi:10.1109/JBHI.2018.2858789.
[32] S. Luo, P. Johnston, A review of electrocardiogram filtering, J. Electrocardiol. 43 (2010) 486–496, doi:10.1016/j.jelectrocard.2010.07.007.
[33] X. Zhai, C. Tin, Automated ECG classification using dual heartbeat coupling based on convolutional neural network, IEEE Access 6 (2018) 27465–27472, doi:10.1109/ACCESS.2018.2833841.
[34] 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.
[35] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Comput. Biol. Med. 100 (2018) 270–278, doi:10.1016/j.compbiomed.2017.09.017.
[36] S.C. Bollepalli, S.S. Challa, S. Jana, Robust heartbeat detection from multimodal data via CNN-based generalizable information fusion, IEEE Trans. Biomed. Eng. (2018) 1–1, doi:10.1109/TBME.2018.2854899.
[37] Z. Lan, O. Sourina, L. Wang, R. Scherer, G.R. Müller-Putz, Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets, IEEE Trans. Cogn. Dev. Syst. (2018) 1–1, doi:10.1109/TCDS.2018.2826840.
[38] S. Chernenko, ECG processing — R-peaks detection [online], (2012). http://www.librow.com/articles/article-13.
[39] M.S. Manikandan, K.P. Soman, A novel method for detecting R-peaks in electrocardiogram (ECG) signal, Biomed. Signal Process. Control. 7 (2012) 118–128, doi:10.1016/j.bspc.2011.03.004.
[40] J. Pan, W.J. Tompkins, A real-time QRS detection algorithm, IEEE Trans. Biomed. Eng. BME-32 (1985) 230–236, doi:10.1109/TBME.1985.325532.
[41] H. Khamis, R. Weiss, Y. Xie, C. Chang, N.H. Lovell, S.J. Redmond, QRS detection algorithm for telehealth electrocardiogram recordings, IEEE Trans. Biomed. Eng. 63 (2016) 1377–1388, doi:10.1109/TBME.2016.2549060.
[42] J.P.V. Madeiro, P.C. Cortez, J.A.L. Marques, C.R.V. Seisdedos, C.R.M.R. Sobrinho, An innovative approach of QRS segmentation based on first-derivative, Hilbert and wavelet transforms, Med. Eng. Phys. 34 (2012) 1236–1246, doi:10.1016/j.medengphy.2011.12.011.
[43] S.D. Greenwald, The development and analysis of a ventricular fibrillation detector Thesis, Massachusetts Institute of Technology, 1986.