Expert Systems With Applications 181 (2021) 115031

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Automated classification of five arrhythmias and normal sinus rhythm based on RR interval signals

Oliver Faust [a,*], U. Rajendra Acharya [b,c,d]

[a] Engineering and Mathematics, Sheffield Hallam University, Sheffield, UK
[b] Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore
[c] Department of Bioinformatics and Medical Engineering, Asia University, Taichung, Taiwan
[d] School of Management and Enterprise University of Southern Queensland, Springfield, Australia

ARTICLE INFO

Keywords:
Computer aided diagnosis
Arrhythmia detection
Deep learning
Residual Neural Network
M-health

ABSTRACT

Arrhythmias are abnormal heart rhythms that can be life-threatening. Atrial Fibrillation (AFIB), Atrial Flutter (AFL), Supraventricular Tachycardia (SVT), Sinus Tachycardia (ST), and Sinus Bradycardia (SB) are common arrhythmias that affect a growing number of patients. In this paper we describe a method to detect these arrhythmias in RR interval signals. We propose a deep learning algorithm to discriminate these fife arrhythmias and Normal Sinus Rhythm (NSR). The deep learning model was trained and tested with data from 10093 subjects. We used 10-fold cross-validation to establish the performance results. The overall accuracy for the six-class problem was 98.37%. When considering the binary problem of arrhythmia versus NSR, where the arrhythmia group is formed by combining the data from all fife arrythmias, the performance results are: Accuracy (ACC) = 98.55%, Sensitivity (SEN) = 99.40%, Specificity (SPE) = 94.30%. These results indicate that it is possible to discriminate RR interval sequences from SVT, ST, SB, AFIB, AFL, and NSR subjects with minimal error. Furthermore, the proposed model can provide a robust and independent second opinion when it comes to a decision if arrhythmia is present or not. Another positive aspect of the proposed arrhythmia detection algorithm is economic viability. RR interval signals are cost-effective to measure, communicate, and process. The discriminate powers of the proposed algorithm together with the advent of wearable technology and m-health infrastructure might lead to pervasive long-term arrhythmia monitoring. The detection results can support early diagnosis which helps to reduce the burden of the disease.

1. Introduction

Patients living with arrhythmia are a growing public health problem (Mozaffarian et al., 2016). The most common arrhythmia is (AFIB) with an estimated prevalence of 1.5% to 2% in industrialized countries (Fuster et al., 2011), affecting approximately 4.5 million in the European Union (Lainscak, Dagres, Filippatos, Anker, & Kremastinos, 2008), and ≈2.2 million people in the United States. The prevalence increases with age (De Caterina, Atar, Hohnloser, & Hindricks, 2012) to six to eight percent in people over 65 years of age (Colilla et al., 2013; Krijthe et al., 2013). It is predicted that the number of AFIB cases will rise 2.5-fold in line with an aging society over the next 50 years (Gillis, Krahn, Skanes, & Nattel, 2013). AFL is another atrial arrhythmia that is less common than AFIB with a prevalence of 0.09% of the population, but whose projected disease burden is also set to double at least by 2050 (Naccarelli, Varker, Lin, & Schulman, 2009). Over half of patients with AFL also have AFIB, however AFL also carries an independently increased risk of heart failure and stroke compared to those with a normal heart rhythm (Fuster et al., 2001). SVT affects an estimated 35 per 100,000 individuals (Colucci, Silver, & Shubrook, 2010; Delacrétaz, 2006). It occurs more commonly in women than men (Fox et al., 2008). In the absence of structural heart disease, SVT can manifest at any age, but it starts to appear most often between the ages of 12 and 30 years (Ferguson & DiMarco, 2003). The prevalence of (ST) in middle-aged subjects is estimated to be around 1.39% (Still et al., 2005). The prevalence of (SB) in the general population is unknown, because most cases are classed as variations of NSR. To find a prevalence indication, we have to look at the sick sinus syndrome, which is an underlying cause of SB. Even for sick sinus syndrome the prevalence in the general population is not known, however in cardiac patients it has been estimated to

* Corresponding author.
E-mail addresses: oliver.faust@gmail.com (O. Faust), aru@np.edu.sg (U.R. Acharya).

https://doi.org/10.1016/j.eswa.2021.115031
Received 1 November 2020; Received in revised form 12 March 2021; Accepted 9 April 2021
Available online 20 April 2021

be 3 in 5000 (Semelka, Gera, & Usman, 2013; Jensen et al., 2014). Increased healthcare costs are associated with the increased risk that arrhythmia patients develop more serious conditions. This increased cost is due to higher hospitalization rates and higher utilization of health resources. Hence, arrythmias pose a great clinical and economic burden on society (Sanoski, 2009). The high prevalence of arrhythmia in the general population is also of concern during the coronavirus disease (COVID-19) pandemic, because studies show that this disease has a detrimental effect on the pathophysiology of cardiac and arrhythmic complications (Kochi, Tagliari, Forleo, Fassini, & Tondo, 2020; Clerkin et al., 2020; Ganatra, Hammond, & Nohria, 2020).

Apart from a direct impact on the cardiac health of patients, coronavirus disease has also exposed weaknesses in arrhythmia healthcare (Tanne et al., 2020). Early diagnosis and adequate treatment monitoring require measurements that can provide objective information about the presence of disease symptoms. However, social distancing and hygiene regulation, introduced in response to the COVID-19 pandemic, have disrupted the data acquisition processes (Böhm, Frey, Giannitsis, Sliwa, & Zeiher, 2020; Driggin et al., 2020). More fundamentally, the pandemic has exposed underlying problems, such as detection quality and test space coverage, for arrhythmia detection technology. The detection methods must be accurate, because a misdiagnosis is likely to have adverse consequences for a patient. A false negative diagnosis might lead to ineffective or no treatment. A false positive diagnosis carries risk due to unnecessary treatment. Detection quality is well addressed with current diagnosis support technology. However, the current technology fails to cover sufficient test space. NHS England has quantified the scale of that failure with an estimate that 21% of arrhythmia cases remain undetected[1]. The fundamental problem is that current technology under samples the test space. For example, heart signals might be measured and analyzed before arrhythmia onset, or when the patient is asymptomatic. In both cases, the resulting data does not contain symptom related information. Hence, the arrhythmia cannot be detected. At the time of measurement, the diagnosis is correct; however, the example scenarios indicate that arrhythmia symptoms will appear as time progresses. That is a problem for current diagnosis support technology, because it is not economical to continue or repeat measurements. The consequence of this shortcoming is a significant under detection of arrhythmia in the general population. Longer measurements from more patients are needed to achieve an adequate test space coverage. Inevitably, this will lead to more data which needs to be communicated, stored, and processed. Data processing is crucial for the viability of that approach because it must provide relevant support for arrhythmia diagnosis.

In this paper, we propose a deep learning data processing algorithm to support arrhythmia diagnosis. The deep learning algorithm took the form of a (ResNet) model which automates the classification of SVT, ST, SB, AFIB, AFL, and NSR based on RR interval signals. We have used benchmark data to train and test the model. The test results show that the ResNet model is able to detect the presence of arrhythmia with an (ACC) of 98.55%, a (SEN) of 99.40%, and a (SPE) of 94.30%. These performance measures provide strong support for the hypothesis that it is possible to detect and indeed discriminate arrhythmias in RR interval signals. Hence, this work can be used as a theoretical foundation for computer-aided arrhythmia diagnosis systems which process RR interval data. Cost efficiency and convenience for the patient are the two main advantages of using RR intervals when compared to standard measurements used to detect arrhythmias. These two advantages have the potential to translate into more and longer measurements which will significantly increase the amount of diagnostically relevant data. That data will cover the test space better which will increase the arrhythmia detection rate. Having a higher detection rate will lead to more treatment which will improve outcomes for patients.

To support our thesis about the efficacy of the proposed arrhythmia detection method, we have structured the remainder of this manuscript as follows. The next section provides some medical background on arrhythmias. We discuss the disease symptoms and the standard measurements which are used for diagnosis. This information is relevant to appreciate the methods introduced in Section 3. In that technical part we focus on the thought processes that gave rise to the processing structure used to train and test the deep learning network. Section 4 provides the performance measurement results for the arrhythmia detection system. These results do not stand in isolation, they were achieved by pushing the envelope of our current understanding of arrhythmia detection with physiological signals. The Discussion section highlights this point by introducing relevant research work and comparing our findings with the established knowledge. Section 6 concludes the paper with final thoughts about the work and its relevance for the medical domain.

2. Background

Arrhythmias are conditions where the heart rhythm is irregular, too slow, or too fast. For most arrhythmias, effective treatment is available. However, this treatment is condition specific, hence it is important to diagnose the individual arrhythmia types accurately. In the remainder of this section we provide some medical background on AFIB, AFL, SVT, ST, and SB. AFIB is not a lethal condition, however it can increase the risk of morbidity and mortality due to disease-related complications, like arterial thrombotic events and cardiac failure (Furberg et al., 1994). Studies show that patients with AFIB have a five-times higher stroke risk (Wolf, Abbott, & Kannel, 1991) and a three-times higher risk of developing heart failure (De Caterina et al., 2012; Friberg et al., 2014). Overall, AFIB doubles the risk of death when compared to healthy people of the same age (Fuster et al., 2011). AFL is another condition which is characterized by an irregular heart rhythm (Sawhney & Feld, 2008). It also confers a thromboembolic risk which elevates the ischemic stroke risk (Ghali, Wasil, Brant, Exner, & Cornuz, 2005). The clinical management of symptoms from AFIB and AFL differs, hence differentiating between them is an important clinical step (Sawhney & Feld, 2008). For instance, while stroke prophylaxis decisions, like whether to anticoagulate a patient, do not differ between AFIB and AFL, treatment decisions for management of symptoms such as palpitations and breathlessness differ greatly. For example, the chances of successful treatment using drug or electrical treatments, such as cardioversion (electrical treatment for restoring normal heart rhythm), being significantly higher in patients with AFL than AFIB (Taha, Reddy, Xue, & Swiryn, 2000).

SVT is characterized by an abnormally fast heart rhythm (Ganz & Friedman, 1995; Camm & Garratt, 1991). For patients with paroxysmal SVT, most or all of their SVT episodes may naturally cease, intervention is required for patients where the SVT persists. State of the art diagnosis is based on physical examination and (ECG) analysis. These methods may reveal damage to the cardiac muscle, and they can be used to identify any other conditional disturbances. Paroxysmal SVT can sometimes be misdiagnosed as anxiety or a panic disorder (Orejarena et al., 1998). Therefore, long term monitoring and careful signal analysis is necessary to reduce the occurrence of such false negatives (Colucci et al., 2010; Delacrétaz, 2006). ST is also characterized by a faster sinus rhythm. In adults, sinus tachycardia is defined as a heart rate greater than 100 beats per minute (Olshansky & Sullivan, 2013).

SB results from a sinus node dysfunction and the main symptom is lower than normal heart rhythm (Eraut & Shaw, 1971). Slow heart rates can be observed during sleep or at times of high vagal tone. When there is significant SB, a slow junctional escape rhythm or a slow atrial rhythm originating from an ectopic focus can be present (Scholar, 2015).

Arrhythmia alters the rhythm with which the heartbeat occurs. Detecting the heartbeat and recording the time when a beat occurred is a first step to determine the heart rhythm. In medical terminology, a

[1] https://www.gov.uk/government/publications/health-matters-preventing-cardiovascular-disease/health-matters-preventing-cardiovascular-disease

heartbeat is known as R peak and the occurrence of that R peak is measured relative to the previous R peak. The time from one heartbeat to the next heartbeat is known as the RR interval. The heart rhythm causes characteristic variations in the RR interval values. Hence, analyzing RR interval signals can be used to determine the heart rhythm which is a good arrhythmia indicator. In the next section we introduce the RR interval analysis methods that lead to the classification of five arrhythmias and NSR.

3. Methods

In this section we introduce the signal processing methods used in the study. As such, these methods were used to create a system to classify five arrythmias and NSR. The system design starts by identifying suitable benchmark data and by selecting an abstract train and test strategy for the deep learning model. The task for the signal processing system is to refine and rearrange the benchmark data, such that it can be used for training and testing the deep learning system. Two additional requirements were data augmentation and data balancing. Data augmentation increases the amount of data, such that the deep learning model can extract disease specific knowledge from the signals (Lashgari, Liang, & Maoz, 2020). Balancing a dataset means to create the same amount of training data for each class (Chawla, Bowyer, Hall, & Kegelmeyer, 2002; Drummond & Holte, 2003; He & Garcia, 2009). A balanced dataset avoids overexposure of the network to one class during training.

Fig. 1 shows an overview block diagram of the data processing system. The processing starts with extracting SVT, ST, SB, AFIB, AFL, and NSR from the (DB). This opens up one processing path for the data from each of the six signal classes. The next step in the class specific processing path is to split the available data into individual folds. The splitting operation yields a number of RR interval blocks for each fold and a row vector which indicates the total number of RR intervals in each fold. The class specific row vectors are combined to form the *nRR* matrix. The Control block takes *nRR* as input and produces both Scramble factor *SF* and Puncturing target *PT*. These two matrices steer the augmentation of the RR interval blocks. The Augmentation block increases the number of RR intervals, such that there is the same number of RR intervals for each class in the same fold. That concludes the class specific data processing, now the data is conditioned such that it can be used to train and test the ResNet model with 10-fold cross-validation. The first step of 10-fold cross-validation is to collect the class specific data and arrange it into 10 training and 10 testing datasets. The training dataset is passed to the Training block where it is used to train a ResNet deep learning model. Once the training is completed, the model is passed to the Testing block which uses the test dataset to evaluate model performance. To be specific, the model was tested with performance measures derived from a confusion matrix and (ROC). The following sections introduce both data and processing steps in more detail.

3.1. Data

The 'ECG data de-noised'[2] Database, depicted on the left side of Fig. 1, sources the benchmark data to train and test the ResNet algorithm. That DB contained 12-lead ECG signals from 10,646 patients. The signal duration was 10 s and the sampling frequency was 500 Hz. The signals were measured at Chapman University and Shaoxing People's Hospital (Shaoxing Hospital Zhejiang University School of Medicine) (Zheng et al., 2020). Each signal was labeled by a cardiologist to indicate one of 11 common rhythms. The labels came in the form of a table which links disease label and ECG signal file name. Based on that table, we selected both arrhythmia and NSR signals. The DB contained 1826 NSR, 445 AFL, 1780 AFIB, 586 SVT, 1568 ST, and 3888 SB signal files. Table 1 shows the number of patients for each signal class and the accumulated (over the individual patient signals within a class) ECG signal duration. As such, the table entries for the ECG duration reflect the fact that all ECG signals had a length of 10 s, i.e. the ECG signal duration is the number of patients times 10 s. The ECG signal from each patient forms one data block. One such ECG data block contains an array of 12×5000 samples, where 12 indicates the number of leads and 5000 is the number of samples captured within 10 s. The term data block is used in the description of subsequent processing steps to denote the data from one patient. The Fold split processing step, described in the next section, extracts the RR interval from the ECG signals. The last row of Table 1 shows the class specific number of RR intervals.

Fig. 2 shows three signals from one subject for each class. The first signal is the 10 s ECG. The RR intervals RR (s), extracted from the ECG signal, are shown in the plot immediately below the ECG. The detrended RR intervals are shown below the RR intervals as RR_DT (s). Both, RR (s) and RR_DT (s) are introduced in the next section.

3.2. Fold split

The Fold split block determines the RR intervals through QRS detection, and it partitions the data into 10 folds. Fig. 3 shows the Fold split block diagram. As such, the functionality is established with a sequence of two processing steps. The QRS detection step extracts the RR intervals from the ECG signals. The subsequent 'Block to fold' step partitions the available data into 10 folds and it provides information about that partitioning in the form of a vector with the number of RR intervals per fold.

QRS detection determines the time location of the heartbeat, known as the R peak, in ECG signals. As such, QRS is the main structural element in ECG. It is caused by ventricular depolarization that occurs when the heart muscle contracts during the heartbeat. Within the QRS complex, the R wave marks the peak, and the time location of that peak represents the time location of the heartbeat. One RR interval is the time from one R peak to the next. We have used the well-known ecg-kit, a MATLAB toolbox for ECG processing (Demski & Soria, 2016), for QRS detection. Within the ecg-kit framework, the wavedetect algorithm was used (Martínez, Almeida, Olmos, Rocha, & Laguna, 2004). The resulting RR interval sequences were saved, such that the block structure was maintained. Table 1 shows the number of RR intervals for each signal class. As such, this step constitutes a significant data reduction. The following example illustrates the data reduction. There were 1826 NSR ECG data blocks which contained 109560000 samples. After QRS detection, there were only 33976 RR intervals. Hence, the compression ratio achieved by the QRS detection step was 3224.6291.

Fig. 2 shows the extracted RR intervals for the example signals. The y-axis scale indicates the RR interval duration, and the x-axis scale indicates the RR interval location, i.e. the time location where the RR interval ends.

The 'Block to fold' processing step follows the principle of parallel to serial conversion. The stream of RR interval blocks is split by storing the individual blocks in one of 10 bins in a round robin fashion. Table 2 shows the properties of the 10 folds for each class. For example, the data from 1826 NSR patients was split into 10 folds with 183, 183, 183, 182, 182, 182, 182 data blocks respectively. The data Augmentation is governed by the number of class specific RR intervals (*nRR*) in each fold. However, to determine the correct augmentation parameters, *nRR* from all the classes is passed to a Control block, which is described in the next section.

3.3. Control

The Control block steers the data Augmentation based on information from the Fold split. It takes *nRR* from each class as input and produces both Scramble factor *SF* and Puncturing target *PT* as output. The Puncturing target specifies the number of data vectors for each class in a

[2] https://figshare.com/collections/ChapmanECG/4560497/2

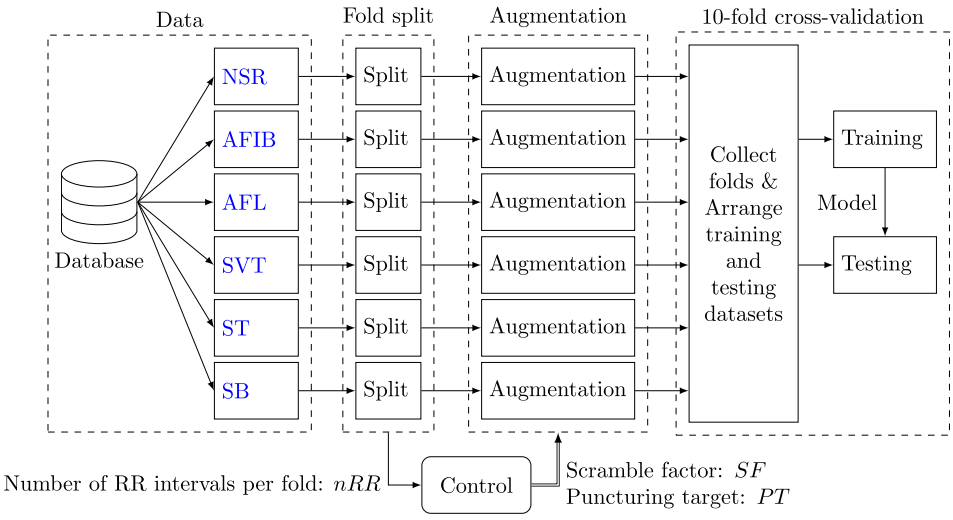

Fig. 1. Overview block diagram.

Table 1
Data properties.

	Class						
Property	SVT	ST	SB	AFIB	AFL	NSR	Total
Number of Patients	586	1568	3888	1780	445	1826	10,093
ECG duration in seconds	5860	15,680	38,880	17,800	4450	18,260	100,930
Number of RR intervals	13,855	24,794	45,087	25,995	7536	19,956	137,223

particular fold. In other words, having that target allows the augmentation block to generate the same number of data vectors for each class in a given fold. Eq. (1) specifies that the puncturing target for a specific fold is the maximum number of data vectors for a class. In this case, the puncturing target is always equal to the number of SB data vectors. For example, the puncturing target for Fold 1 is 4518.

$$PT_{1\times 10} = \max(nRR_{6\times 10}) \tag{1}$$

where max(.) returns a row vector containing the maximum value of each column. The scramble factor indicates how much additional data the Augmentation block needs to generate to meet the puncturing target. As such, the scrambling factor is an integer multiplier which when applied to the number of class specific data vectors in a Fold results in a number equal or larger than the puncturing target.

$$SF_{6\times 10} = \lceil PT_{1\times 10} \oslash nRR_{6\times 10} \rceil \tag{2}$$

where $\oslash$ is the Hadamard division and $\lceil . \rceil$ is the round up operation.

3.4. Augmentation

We have used augmentation to increase the data volume and to balance the dataset. A sequence of Block scrambling, Detrending, Windowing, and Puncturing was used to achieve that functionality. The block diagram, shown in Fig. 4, depicts both Augmentation process structure and data flow.

Block scrambling is used for coarse grain adjustment of the class specific imbalances. The patient scrambling concept is based on the fact that the fold generation algorithm uses the order in which the RR interval block appears in the dataset to establish the fold data. This order impacts on the data vectors, which were created through round robin windowing, described below, because the window length is longer than the amount of RR intervals in any data block. Each data vector contains 100 detrended RR intervals from different patients, as outlined in the next section. Hence, a different sequence of patients in the fold will result in different vectors after the windowing. In the scrambling step, we use this property to generate more data. The scramble factor determines the number of fold specific permutations for a given signal class. For example, Table 2 shows that the AFL data blocks are permutated seven times for fold 1 and six times for the remaining folds.

Detrending removes the DC offset from RR interval signals. Applying that processing method benefits the deep learning step by reducing both required network complexity and training time (Faust, Barika, Shenfield, Ciaccio, & Acharya, 2020). For our study, we have used the detrending and low-pass filter proposed by Fisher, Eleuteri, Groves, and Dewhurst (2012). The filter combination is based on an Ornstein–Uhlenbeck third-order Gaussian process which acts on the RR-interval signal directly (Clifford, Azuaje, & Mcsharry, 2006; Laguna, Moody, & Mark, 1998). After detrending, the datasets contain RR_DT samples. Table 1 shows the number of RR_DT samples. As such, the detrending step does not change the amount of data, hence the number of DD_DT samples is the same as the RR intervals. Fig. 2 shows the detrended version of the RR (s) signal for each signal class. The signal graphs show the effect of removing the DC bias with the Detrending step.

Round robin Windowing augments the data by generating one data vector with 100 elements for each RR_DT sample. That method increases the data volume 100-fold. The vectors were generated by subjecting the class specific data for each fold to a window a length of 100 samples. This window was slid over the RR_DT signal one sample at a time. Round robin refers to the fact that the first 100 RR_DT samples for each dataset were copied at the end, before applying the window. That extension allowed us to create one data vector for each RR_DT sample.

After windowing, we have used puncturing to adjust the data size (Johnson & Khoshgoftaar, 2019). The puncturing algorithm removes equidistant data vectors. This technique ensures that the number of training data, for each of the six classes in a fold, is equal. Table 2 shows the number of RR intervals for a given signal class in each fold. That number is equal to the number of signal vectors after windowing. The puncturing target, defined in Eq. (1), determines how many data vectors are removed. Table 2 shows that for SB the number of data vectors

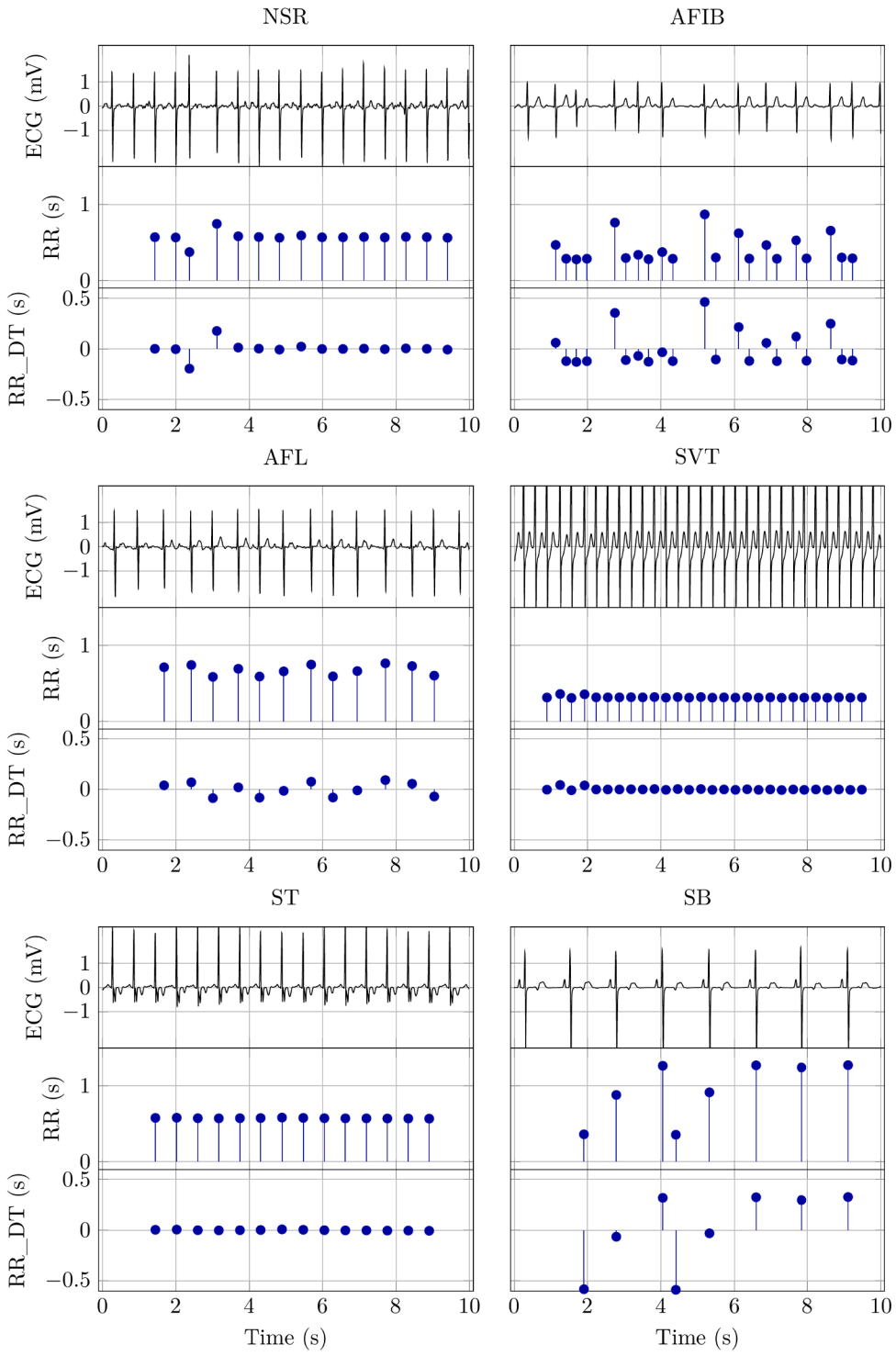

Fig. 2. Example plots from NSR, AFIB, AFL, SVT, ST, and SB signal classes. The ECG signal was measured with the 'V3' lead. The RR-intervals, plotted as RR-intervals over time, were derived from the ECG via QRS detection. The detrended RR-intervals were plotted as RR_DT over time.

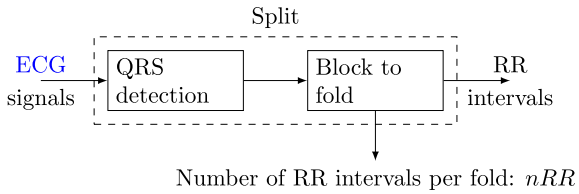

Fig. 3. Fold split block diagram.

(number of RR after scrambling) is equal to the number of data vectors after puncturing, because SB had the highest number of RR intervals in each fold. The number of data vectors after puncturing is the within each fold regardless of the data class. For example, there are 4518 data vectors in fold 1 for each of the six classes. The continuum of data vectors in each fold can be used to train and test the ResNet model with 10-fold cross-validation.

Table 2
Class specific fold properties.

Class	Property	Fold									
		1	2	3	4	5	6	7	8	9	10
NSR	Number of Patients	183	183	183	183	183	183	182	182	182	182
	Number of RR intervals	2015	2025	2027	2026	2025	2024	2026	2026	2024	2028
	Scramble factor	3	3	3	3	3	3	3	3	3	3
	No RR after scrambling	6045	6075	6081	6078	6075	6072	6078	6078	6072	6084
	No vectors after puncturing	4518	4525	4535	4527	4532	4527	4526	4531	4526	4528
AFIB	Number of Patients	178	178	178	178	178	178	178	178	178	178
	Number of RR intervals	2651	2664	2666	2666	2663	2661	2663	2661	2662	2664
	Scramble factor	2	2	2	2	2	2	2	2	2	2
	No RR after scrambling	5302	5328	5332	5332	5326	5322	5326	5322	5324	5328
	No vectors after puncturing	4518	4525	4535	4527	4532	4527	4526	4531	4526	4528
AFL	Number of Patients	45	45	45	45	45	44	44	44	44	44
	Number of RR intervals	742	759	757	759	760	765	758	758	756	764
	Scramble factor	7	6	6	6	6	6	6	6	6	6
	No RR after scrambling	5194	4554	4542	4554	4560	4590	4548	4548	4536	4584
	No vectors after puncturing	4518	4525	4535	4527	4532	4527	4526	4531	4526	4528
SVT	Number of Patients	55	55	55	55	55	55	54	53	54	54
	Number of RR intervals	1344	1372	1369	1373	1364	1365	1369	1367	1366	1369
	Scramble factor	4	4	4	4	4	4	4	4	4	4
	No RR after scrambling	5376	5488	5476	5492	5456	5460	5476	5468	5464	5476
	No vectors after puncturing	4518	4525	4535	4527	4532	4527	4526	4531	4526	4528
ST	Number of Patients	157	157	157	157	157	157	157	157	156	156
	Number of RR intervals	2464	2479	2478	2480	2480	2478	2484	2480	2479	2479
	Scramble factor	2	2	2	2	2	2	2	2	2	2
	No RR after scrambling	4928	4958	4956	4960	4960	4956	4968	4960	4958	4958
	No vectors after puncturing	4518	4525	4535	4527	4532	4527	4526	4531	4526	4528
SB	Number of Patients	389	389	389	389	389	389	389	389	388	388
	Number of RR intervals	4518	4525	4535	4527	4532	4527	4526	4531	4526	4528
	Scramble factor	1	1	1	1	1	1	1	1	1	1
	No RR after scrambling	4518	4525	4535	4527	4532	4527	4526	4531	4526	4528
	No vectors after puncturing	4518	4525	4535	4527	4532	4527	4526	4531	4526	4528

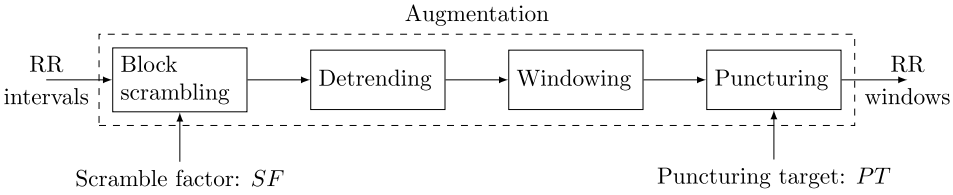

Fig. 4. Augmentation block diagram.

3.5. ResNet 10-fold cross-validation

Overfitting is the main problem for physiological signal classification with deep learning. The term refers to the fact that the deep learning network is able to memorize the signals itself, rather than the signal properties that indicate disease symptoms. In practice, overfitting occurs when the model classifies training data correctly but fails to do so with testing data. There are a range of techniques to avoid or at least reduce overfitting. Model selection plays an important role in that process. For this study, we followed the findings by Ismail Fawaz, Forestier, Weber, Idoumghar, and Muller (2019). In their review on deep learning for time series classification they found that ResNet outperforms all the other tested deep learning models. Fig. 5 shows the data flow structure used to establish the ResNet model. The data flow diagram is composed from standard components which have a direct correspondence in the Python API Keras (Chollet, 2015) which abstracts the Deep learning framework tensorflow (Abadi et al., 2015). The data flow structure shows three shortcut connections which allow information to skip the processing block. Such a structure is known as residual block. This structure can be used to address another limitation of deep learning models, namely the vanishing/exploding gradient problem (Hanin, 2018). From a practical perspective, this problem occurs when more network layers result in lower training accuracy and therefore this problem category is distinct from the general overfitting problem.

3.6. Result analysis methods

The result analysis starts with establishing the confusion matrix based on the validation results. Table 3 defines the confusion matrix in terms of the number of beats with a true and a predicted label: $c_{\text{predictedlabel,truelabel}}$. The predicted label was established with the ResNet classification model. Combining the 6 predicted and 6 true labels results in a confusion matrix with $6 \times 6 = 36$ beat labels. Table 3 shows the arrangement of these beat labels in a 6 by 6 confusion matrix C.

Based on the confusion matrix we establish four different performance measures, namely ACC_{Fold}, ACC_{cl}, SEN_{cl}, and SPE_{cl}. The overall or total accuracy ACC_{Fold} describes the classification performance over the continuum of classes in a specific fold. That performance measure is established by dividing the number of correctly predicted data vectors with the total number of data vectors. Based on the confusion matrix, that task simplifies to summing up all elements of the diagonal and dividing the result by the sum of all elements, as defined in the following equation.

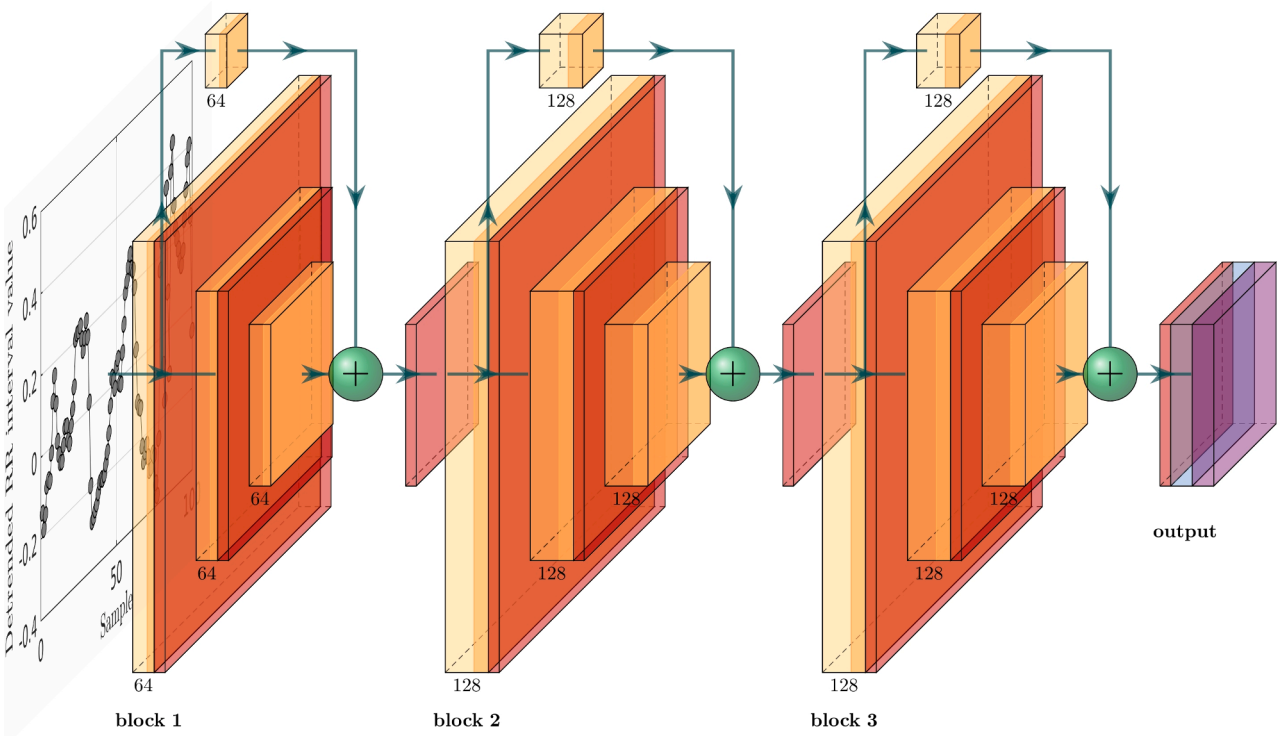

Fig. 5. Network diagram. The data flow starts from the input vector with 100 detrended RR intervals. Block 1 consists of three 1D convolution layers with 64 filters each and the kernel size is 8, 5, and 3 respectively. Between each convolution layers there are batch normalization and activation layers. The residual path is provided by a 1D convolution (64 filters and a kernel size of 1) with batch normalization. Blocks 2 and 3 have the same setup, the number of filters in the convolution layers is 128. The output is established with global average pooling and a dense layer with sigmoid activation.

Table 3
Confusion matrix C for the 6-class problem. Each element of the confusion matrix is labeled as $c_{\text{Truelabel,Predictedlabel}}$ where Truelabel, Predictedlabel $\in \{SVT, ST, SB, AFIB, AFL, NSR\}$.

		Predicted label					
		SVT	ST	SB	AFIB	AFL	NSR
True label	SVT	$c_{SVT,SVT}$	$c_{SVT,ST}$	$c_{SVT,SB}$	$c_{SVT,AFIB}$	$c_{SVT,AFL}$	$c_{SVT,NSR}$
	ST	$c_{ST,SVT}$	$c_{ST,ST}$	$c_{ST,SB}$	$c_{ST,AFIB}$	$c_{ST,AFL}$	$c_{ST,NSR}$
	SB	$c_{SB,SVT}$	$c_{SB,ST}$	$c_{SB,SB}$	$c_{SB,AFIB}$	$c_{SB,AFL}$	$c_{SB,NSR}$
	AFIB	$c_{AFIB,SVT}$	$c_{AFIB,ST}$	$c_{AFIB,SB}$	$c_{AFIB,AFIB}$	$c_{AFIB,AFL}$	$c_{AFIB,NSR}$
	AFL	$c_{AFL,SVT}$	$c_{AFL,ST}$	$c_{AFL,SB}$	$c_{AFL,AFIB}$	$c_{AFL,AFL}$	$c_{AFL,NSR}$
	NSR	$c_{NSR,SVT}$	$c_{NST,SVT}$	$c_{NSR,SB}$	$c_{NSR,AFIB}$	$c_{NST,AFL}$	$c_{NSR,NSR}$

$$ACC_{Fold} = \frac{\sum_{i \in Class\ set} c_{i,i}}{\sum_{i,j \in Class\ set} c_{i,j}} \tag{3}$$

Having defined a measure that looks at the overall classification performance of the ResNet model, we turn our attention to class specific performance measures. These measures help us to compare the current study with other studies that had a different setup. To be specific, we investigate the ResNet classification performance for a specific class *cl* versus the remainder of the classes. In effect, that reduces the six-class problem to a binary problem – one binary problem for each class. For this class specific binary problem there are two ways in which the classification can be correct and two ways in which it can be wrong. We define the correct cases as (TP), (TN), and the wrong cases as (FP), (FN). Positive indicates a data vector belongs to the selected class *cl* and negative indicates a data vector does not belong to that class. The following equations provide the mathematical definitions:

$$TP_{cl} = c_{cl,cl} \tag{4}$$

$$TP_{cl} = \left(\sum_{i \in Class\ set} c_{i,i} \right) - c_{cl,cl} \tag{5}$$

$$FP_{cl} = \left(\sum_{i \in Class\ set} c_{i,cl} \right) - c_{cl,cl} \tag{6}$$

$$FN_{cl} = \left(\sum_{i \in Class\ set} c_{cl,i} \right) - c_{cl,cl} \tag{7}$$

where $cl \in \{Class\ set\}$ and Class set = {SVT, SB, ST, AFIB, AFL, NSR}.

These definitions were used to establish the performance measures of ACC, SPE, and SEN for the individual class:

$$ACC_{cl} = \frac{TP_{cl} + TN_{cl}}{TP_{cl} + TN_{cl} + FP_{cl} + FN_{cl}} \tag{8}$$

$$SEN_{cl} = \frac{TP_{cl}}{TP_{cl} + FN_{cl}} \tag{9}$$

$$SPE_{cl} = \frac{TN_{cl}}{TN_{cl} + FP_{cl}} \tag{10}$$

From a medical perspective it is important to determine if a patient has arrhythmia or not. This is a two-class problem which is related to the individual performance measures for NSR. To establish these individual performance measures, we considered the NSR data vector as positive and all data vectors with an arrhythmia label as negative. However, for a medical discussion it is more appropriate to consider the normal case NSR as negative and the diseased case (arrhythmia) as positive. Hence, for the arrhythmia no-arrhythmia case True and False swap in Eqs. (8)–(10). As a consequence, the accuracy values stay the same, i.e. $ACC_{NSR} = ACC_{\text{arrithmia/no-arrithmia}}$. However, the values for sensitivity and specificity swap, i.e. $SEN_{NSR} = SPE_{\text{arrithmia/no-arrithmia}}$ and $SPE_{NSR} = SEN_{\text{arrithmia/no-arrithmia}}$.

A ROC curve illustrates how the threshold level influences the diagnostic ability of a binary classifier (Fawcett, 2006). The area under the ROC curve indicates the general performance of the classifier, i.e. an area closer to 1 indicates a better classification performance. The class specific Eqs. (9) and (10) were used to calculate True Positive Rate and False Positive Rate, respectively. The micro-average is the mean of the individual results for the three classes. The macro-average is calculated by aggregating all False Positive Rates[3].

4. Results

The results, reported in this section, are based on 10-fold cross-validation. Table 4 details the ACC_{Fold} and validation loss results achieved during cross validation. The ACC_{ALL} of 98.37% indicates the overall accuracy for the six-class problem. Fig. 6 shows the ACC_{Fold} and validation loss results over the training epochs. As such there is not much movement in the graph. That means the ResNet was able to capture sufficient knowledge to achieve high ACC_{Fold} and low validation loss early in the training process.

With the methods, outlined in Section 3.6, we have generated one confusion matrix C_{Fold} for each test fold. Based on these confusion matrices, we calculated the class specific performance measures. Table 5 details the confusion matrices for each fold and the class specific performance results. The last six rows were labeled as Fold Test All to indicate the performance measures over all folds. These measures were based on the confusion matrix C_{All} which was established with the element wise sum of the individual confusion matrices:

$$C_{\text{All}} = \sum_{Fold=1}^{10} C_{Fold} \tag{11}$$

Based on the reasoning about detecting arrhythmia and no-arrhythmia we use the performance measures, stated in the last row of Table 5 as:

- $ACC_{\text{arrithmia/no-arrithmia}} = 98.55\%$
- $SEN_{\text{arrithmia/no-arrithmia}} = 99.40\%$
- $SPE_{\text{arrithmia/no-arrithmia}} = 94.30\%$

The ROC curve, shown in Fig. 7, provides a graphical representation of the classification results. The large area under the curve is a direct result of the good classification performance indicated by the performance measures stated in Table 5.

5. Discussion

The proposed arrhythmia classification method does not stand in isolation. Over the years a substantial body of literature has grown which methods for arrhythmia detection. When discussing history, we have to begin somewhere, and we chose 1983 as starting point. That was the year when (Moody & Mark, 1983) proposed a new method for detecting AFIB using RR intervals. They employed a Hidden Markov model for decision making and hand optimized features. That approach has set the blueprint for a wide range of scientific work on arrithmia detection (Castillo, Melin, Ramírez, & Soria, 2012; Melin, Miramontes, & Prado-Arechiga, 2018; Ramirez, Melin, & Prado-Arechiga, 2019). By now, this approach is outdated, because of the limitations of both feature engineering and choice of classification method. Classical machine classifiers, such as the Hidden Markov model, fail to generalize findings from a limited training dataset (Faust, Hagiwara, Hong, Lih, & Acharya, 2018). One abstract explanation for that failure is that classical machine classifier lack complexity due to their rules-based nature. For example, a Hidden Markov model is based on states, observations, and transition probabilities. Such a structure is designed with some sort of understanding on what a specific state transition means. This understanding, or indeed this assumption might not hold when it comes to a large volume of unknown data. Furthermore, this design limitation is also the root cause for the failure of these decision structures to deal with high dimensional input. To be specific, classical machine learning operates based on low dimensional input or feature vectors. For example, the method proposed by Moody and Mark used around 20 RR intervals to determine three state transitions. Signal processing methods, such as filtering, were used to extract features which improve the state transition accuracy. However, these signal processing methods limit the amount of information for the decision-making system. Furthermore, which feature extraction methods are used is up to the designer and resolving that choice might lead to sub-optimal solutions. In contrast, deep learning systems do not suffer from these limitations. Design decisions, such as model selection and hyper-parameter tuning happen on a much higher level of abstraction. Furthermore, all design decisions are validated, i.e. we check if a specific decision improves the classification performance. In the past, statistical feature ranking was used to determine which features to use. Unfortunately, all the best features might measure the same information and selecting the best features might lead to selecting the same information. To quantify the differences between the different approaches, it is necessary to have benchmark datasets. Currently, lots of arrhythmia research is based on the atrial fibrillation database (afdb) (Moody & Mark, 2001). This dataset was first used by Moody and Mark in their study (Moody & Mark, 1983) before it was published on Physionet (Goldberger et al., 2000).

The availability of benchmark datasets opened the arrhythmia detection problem to a wider range of researchers. Especially the introduction of deep learning to time series classification attracted many experts to study that problem. This interest sparked a significant amount of scientific work which is documented by a large number of publications (Faust, Ciaccio, & Acharya, 2020). Yildirim, Talo, Ciaccio, San Tan, and Acharya (2020) were the first to build an automated arrhythmia detection system based on ECG signals from the ecgdb, described in Section 3.1. Our study was the first to use that DB for RR interval-based arrhythmia detection. The lower data rate of RR interval signals rendered some signal data in the ecgdb data base impractical to process. To be specific, Yildirim et al. (2020) included (AVNRT), (AVRT), (SI), and (SAAWR) in their ecgdb study, however the number of 10 s ECG for these rhythms is 16, 87, 397, 121, 7 respectively. In the current study these rhythms were excluded because the short ECG traces resulted in an insufficient number of RR intervals to train the ResNet effectively. Ivanovic, Atanasoski, Shvilkin, Hadzievski, and Maluckov (2019) advanced the science of RR interval analysis with an automated classification system for AFIB, AFL, and NSR. In our work, we have extended the number of detectable arrhythmia from two to five by including SVT, ST, and SB as signal classes. Furthermore, we could improve the classification accuracy from 88% to 98.55%.

Table 6 provides a summary of selected work on automated arrhythmia detection based on RR interval and ECG signals. The table

[3] https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html

Table 4
ACC_{Fold} and validation loss for each fold as well as for all folds. The validation loss over all folds was established by averaging the individual loss results.

	Fold										
	1	2	3	4	5	6	7	8	9	10	All
ACC_{Fold}	99.35	97.14	98.14	98.06	99.17	98.59	98.33	97.52	98.50	98.84	**98.37**
Validation loss	0.019	0.137	0.103	0.075	0.043	0.059	0.060	0.130	0.086	0.059	0.077

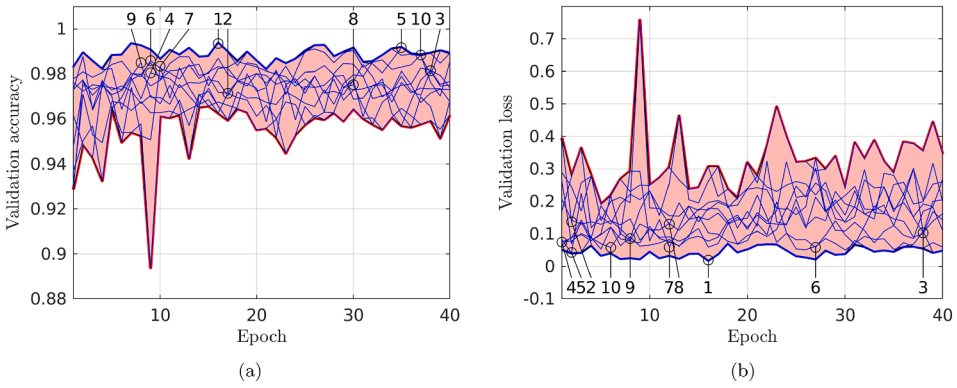

Fig. 6. Validation performance for each test fold over 40 epochs. (a) Validation ACC over epochs. The markers 'o' indicate the maximum validation accuracy for an individual fold. The highlighted area is spanned by maximum and minimum validation ACC from the continuum of test folds for each epoch. (b) Validation loss over epochs. The markers 'o' indicate the minimum validation loss for an individual fold. The highlighted area is spanned by maximum and minimum validation loss from the continuum of test folds for each epoch.

details the signal processing methods used to establish the system which trains and tests the classification algorithm. The next three columns indicate (a) data type, (b) DB used, and (c) rhythm to be detected. For this table, we selected only studies that use either RR intervals or ECG to ensure a good comparability of the study outcomes. Having restricted the data type also places a constraint on the DBs that can be used for such studies. The DB column indicates that Physionet is a good data source together with the ecgdb, described in Section 3.1. Having data from multiple DBs in one study increases the processing complexity and it impacts on the classification performance, because of the inevitable differences between measurement setups. Usually, the signals within one DB were captured with the same measurement setup having a specific number of leads and the signals are converted from analogue to digital with the same electronic chip. However, it is unlikely that these parameters are the same across DBs. For example, signals from the ltafdb were sampled with 128 Hz whereas signals from afdb were sampled with 250 Hz and the signals used in this study (ecgdb) were sampled with 500 Hz. The number of rhythms constitutes the number of classes, when the problem is posed in a machine learning context. In general, the level of difficulty for classifier increases with the number of classes. That impacts on the performance as expressed by ACC, SPE, and SEN. In the table, we report the performance measures for the arrhythmia no-arrhythmia problem. This was done to improve the study comparability. Apart from these general considerations, Table 6 shows that seven studies include (VFIB) as one distinct abnormal rhythm. As such, VFIB can lead to sudden cardiac death which makes it a valid subject of study. This condition was excluded from the current study, because of technical reasons including data availability and study setup.

Reporting only the performance on the arrhythmia no-arrhythmia problem puts studies that use a wider range of rhythms at a disadvantage. The following reasoning shows the extend of the disadvantage. A binary classifier can achieve a $100\%/2 = 50\%$ accuracy by generating a random label for each test input. With the same method, a 6-class classifier achieves only $100\%/6 \approx 16.67\%$. The confusion matrix for all test folds C_{ALL}, shown in Table 5, shows that the ResNet classifier has difficulties to discriminate NSR. Hence, the performance on the arrhythmia no-arrhythmia problem is inferior when compared to three and four class problems. Especially, the SPE of 89.18%[4] leaves room for improvement.

Table 6 shows that the proposed ResNet deep learning model achieves the highest classification accuracy for all studies based on RR intervals signals. To the best of our knowledge, our study was the first to employ this type of deep learning model for arrhythmia detection. The second part of Table 6 shows that our results are similar to ECG based studies. Indeed only (Pudukotai Dinakarrao & Jantsch, 2018; Xia, Wulan, Wang, & Zhang, 2018) report higher accuracy scores. However, both studies posed a two-class problem, whereas our study took six-classes into consideration. Being trained with more signal classes implies that our model has seen a wider range of symptoms which should make it more robust. Therefore it is expected that the ResNet will perform well in a practical setting with unknown input data. The fact that RR intervals were used to achieve these results is a significant achievement, because such signals are easier to measure, communicate, store and process than ECG signals. The following list summarizes the original contribution of this work:

- Highest arrhythmia detection accuracy for studies based on RR intervals.
- Highest accuracy for studies that considered more than two classes.
- Accuracy values similar to ECG based studies.
- Six-class problem investigated.
- ResNet with a low computational complexity for inference was used.

5.1. Limitations

Inevitably for this type of study there are limitations that stem from choices during the study setup. In our case, the study setup started with the recognition that appropriate benchmark data is available. The main task during the design phase was to choose signal processing methods that could be used to translate the available data into a format which is compatible with the machine classification algorithm. As such, the chosen ResNet deep learning algorithm required us to execute extensive data augmentation. The hidden assumption was that in a practical setting the data format is likely to change, but the requirements posed by the decision support will stay the same. To be specific, it is unlikely to encounter snippets of 10 s ECG in a practical setting. This fundamental limitation creates some uncertainty on whether the study results can be repeated in a practical setting. That calls for more research validating

[4] SPE for the arrhythmia no-arrhythmia problem

Table 5
Class specific results for each individual test fold and over All folds. *Fold* indicates the test fold with which the ResNet algorithm was tested.

Fold	*cl*	ACC_{cl} (%)	SEN_{cl} (%)	SPE_{cl} (%)	Test Fold specific confusion matrix: C_{Fold}					
1	SVT	99.83	100.00	99.80	4518	0	0	0	0	0
	ST	99.97	100.00	99.96	0	4518	0	0	0	0
	SB	100.00	100.00	100.00	0	0	4518	0	0	0
	AFIB	99.96	99.76	100.00	3	8	0	4507	0	0
	AFL	99.56	98.69	99.73	0	0	0	0	4459	59
	NSR	99.39	97.68	99.74	43	0	0	1	61	4413
2	SVT	99.49	99.62	99.46	4478	0	0	0	0	17
	ST	98.56	92.10	99.87	117	4140	0	4	0	234
	SB	100.00	99.98	100.00	0	0	4494	0	1	0
	AFIB	99.98	99.98	99.98	0	0	0	4494	0	1
	AFL	98.61	97.22	98.89	0	0	0	0	4370	125
	NSR	97.58	93.93	98.31	0	29	0	0	244	4222
3	SVT	99.62	99.04	99.74	4414	41	0	0	0	2
	ST	99.57	99.55	99.57	18	4437	0	0	0	2
	SB	100.00	100.00	100.00	0	0	4457	0	0	0
	AFIB	99.84	99.08	100.00	0	41	0	4416	0	0
	AFL	98.72	96.48	99.17	0	0	0	0	4300	157
	NSR	98.51	94.70	99.27	40	12	0	0	184	4221
4	SVT	99.88	100.00	99.85	4508	0	0	0	0	0
	ST	99.87	99.22	100.00	32	4473	0	0	0	3
	SB	99.93	99.58	100.00	0	0	4489	0	0	19
	AFIB	100.00	100.00	100.00	0	0	0	4508	0	0
	AFL	98.25	96.47	98.61	0	0	0	0	4349	159
	NSR	98.17	93.06	99.20	0	0	0	0	313	4195
5	SVT	99.72	98.65	99.94	4589	16	0	0	0	47
	ST	99.78	99.33	99.87	2	4621	0	0	0	29
	SB	100.00	100.00	100.00	0	0	4652	0	0	0
	AFIB	100.00	99.98	100.00	0	0	1	4651	0	0
	AFL	99.60	99.89	99.54	0	0	0	0	4647	5
	NSR	99.24	97.16	99.65	12	14	0	0	106	4520
6	SVT	99.78	99.82	99.77	4544	7	0	0	0	1
	ST	99.69	98.88	99.86	50	4501	0	0	0	1
	SB	100.00	100.00	100.00	0	0	4552	0	0	0
	AFIB	99.90	100.00	99.88	0	0	0	4552	0	0
	AFL	98.99	97.23	99.35	0	0	0	0	4426	126
	NSR	98.79	95.58	99.44	2	25	0	26	148	4351
7	SVT	99.82	99.76	99.84	4531	1	0	0	0	10
	ST	99.76	99.58	99.80	19	4523	0	0	0	0
	SB	99.90	99.38	100.00	0	0	4514	28	0	0
	AFIB	99.70	100.00	99.64	0	0	0	4542	0	0
	AFL	98.95	98.61	99.02	0	0	0	1	4479	62
	NSR	98.51	92.67	99.68	17	44	0	52	220	4209
8	SVT	99.10	95.33	99.87	4183	43	0	0	0	162
	ST	99.54	99.02	99.64	28	4345	0	0	0	15
	SB	99.92	100.00	99.91	0	0	4388	0	0	0
	AFIB	99.89	99.38	99.99	0	6	20	4361	1	0
	AFL	98.66	99.77	98.43	0	0	0	0	4378	10
	NSR	97.88	91.61	99.14	0	27	0	2	339	4020
9	SVT	99.92	99.96	99.91	4540	0	0	0	0	2
	ST	99.85	99.14	100.00	0	4503	0	0	0	39
	SB	100.00	100.00	100.00	0	0	4542	0	0	0
	AFIB	100.00	100.00	100.00	0	0	0	4542	0	0
	AFL	98.72	98.52	98.75	0	0	0	0	4475	67
	NSR	98.50	93.35	99.52	20	0	0	0	282	4240
10	SVT	99.60	99.98	99.53	4432	0	0	0	0	1
	ST	99.37	99.98	99.24	1	4432	0	0	0	0
	SB	100.00	100.00	100.00	0	0	4433	0	0	0
	AFIB	99.99	99.95	100.00	0	0	0	4431	0	2
	AFL	99.87	100.00	99.84	0	0	0	0	4433	0
	NSR	98.84	93.12	99.99	103	167	0	0	35	4128
All	SVT	99.68	99.22	99.77	44737	108	0	0	0	242
	ST	99.60	98.68	99.78	267	44493	0	4	0	323
	SB	99.97	99.89	99.99	0	0	45039	28	1	19
	AFIB	99.93	99.82	99.95	3	55	21	45004	1	3
	AFL	98.99	98.29	99.14	0	0	0	1	44316	770
	NSR	98.55	94.30	99.40	237	318	0	81	1932	42519

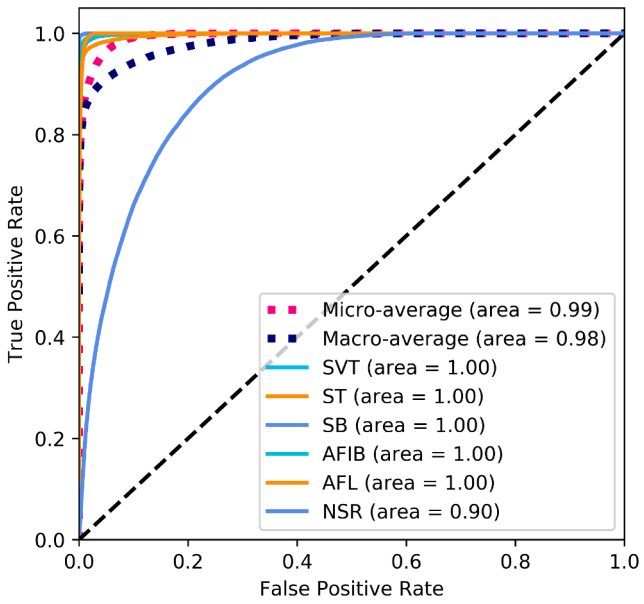

Fig. 7. ROC curve based on the results from all 10 folds.

the ResNet model with more data in a clinical setting.

Another limitation comes from the fact that we extracted RR intervals from ECG data. In a practical setting, a sensor might use a different method to establish the RR intervals. To be specific, there is a wide range of algorithms in existence which extract RR intervals from ECG signals. As such, we have chosen wavedetect, see Section 3.2. More research is needed to establish whether different methods of extracting the RR interval require different classification models.

5.2. Future work

In future we plan to address the comparably low SPE by focusing on the NSR cases during training and thereby optimize the training results within the given training iterations (epochs) (Jiang et al., 2019). Another strategy to address the problem might be to use more data from a wider range of subjects. That is likely to require sourcing the additional NSR data from different DBs, such as Physionet's nsrdb (Goldberger et al., 2000), and fantasia (Iyengar, Peng, Morin, Goldberger, & Lipsitz, 1996). Extending the study setup by sourcing data from a wider range of DBs would allow to include more classes. A reasonable choice for such an extension would be VFIB, because of the medical need to detect this condition (Chugh et al., 2008). Furthermore, Table 6 shows that this signal class was included in at least seven studies. Hence, adding this signal class would also improve comparability and thereby foster competition with the clear goal to improve the arrhythmia detection quality.

Using RR interval measurements to support arrhythmia detection can have a significant impact on the clinical diagnosis process, because the signals can be acquired in the patient environment and communicated in real time to a medical facility. The real time nature of such an approach opens up the possibility for shorter response times, longer observation duration, and continuous patient compliance monitoring (Faust, Yu, & Acharya, 2015). However, to realize the benefits, we must find a safe, reliable, and functional way for human experts to work together with machine algorithms (Faust, Acharya, & Tamura, 2012). In future, that cooperative environment might take the form of a hybrid decision support platform for arrhythmia monitoring (Lei et al., 2021; Faust, Lei, Chew, Ciaccio, & Acharya, 2020).

6. Conclusion

In this paper we present the signal processing methods which help us to automate the classification of SVT, ST, SB, AFIB, AFL, and NSR RR

Table 6
Selected arrhythmia detection studies using RR-intervals and ECG. DBs used were: MIT-BIH Atrial Fibrillation Database (afdb), MIT-BIH Arrhythmia Database (mitdb), MIT-BIH Malignant Ventricular Arrhythmia Database (vfdb), Creighton University Ventricular Tachyarrhythmia Database (cudb), MIT-BIH Normal Sinus Rhythm Database (nsrdb), MIT-BIH Long Term Database (ltdb), European ST-T Database (edb), and ecgdb. Hospital data comes from non-publicly accessible databases.

Author year	Method	Data		Performance in %		
		DB	Rhythm	ACC	SPE	SEN
RR interval based studies						
Current	Detrending, ResNet	ecgdb	SVT, ST, SB, AFIB, AFL, NSR	**98.55**	**94.30**	**99.40**
Ivanovic et al. (2019)	CNN, LSTM	Hospital	NSR, AFIB AFL	88		87.09
Faust et al. (2018)	LSTM	afdb	AFIB NSR	98.39	98.32	98.51
Henzel et al. (2017)	Statistical features with generalized Linear Model	afdb	AFIB NSR	93	95	90
Petrėnas et al. (2015)	Median filter with threshold	nsrdb, afdb	AFIB NSR		98.3	97.1
Zhou et al. (2014)	Median filter & Shannon entropy with threshold	ltafdb, afdb, nsrdb	AFIB NSR	96.05	95.07	96.72
Moody and Mark (1983)	Threshold, Hidden Markov model	afdb	AFIB, NSR			96.15
ECG based studies						
Yildirim et al. (2020)	LSTM	ecgdb	AVRT, SI, SVT, SAAWR, ST, SB, AFIB, AFL, NSR	96.13		
Fujita and Cimr (2019)	CNN with normalization	afdb, mitdb, vfdb	AFIB, AFL, VFIB, NSR	98.45	99.87	99.27
Pudukotai Dinakarrao and Jantsch (2018)	Daubechies-6 with counters Anomaly detector	mitdb	AFIB, VFIB	99.19	98.25	78.70
Salem et al. (2018)	Spectogram with CNN	afdb nsrdb vfdb edb	AFIB, AFL VFIB NSR	97.23		
Xia et al. (2018)	STFT/SWT with CNN	afdb	AFIB	98.63	98.79	97.87
Acharya et al. (2017)	CNN with Z-score	afdb, mitdb, vfdb	AFIB, AFL, VFIB, NSR	92.50	98.09	93.13
Desai et al. (2016)	RQA with DecisionTree, RandomForest, RotationForest	afdb, mitdb, vfdb	AFIB, AFL, VFIB, NSR	98.37		
Acharya et al., 2016	Thirteen nonlinear features with ANOVA with KNN and DT	afdb, mitdb, vfdb	AFIB, AFL, VFIB, NSR	97.78	99.76	98.82
Hamed and Owis (2016)	DWT, PCA and SVM	afdb	AFIB, AFL, NSR	98.43	96.89	98.96
Yuan et al. (2016)	Unsupervised autoencoder NN Softmax regression	afdb, nsrdb, ltdb, hospital	AFIB	98.18	98.22	98.11
Muthuchudar and Baboo (2013)	UWT NN	afdb	AFIB, VFIB, NSR	96		

interval signals. We carefully arranged the available benchmark data into ten folds for training and testing a deep learning model. To be specific, the data was arranged such that data from a specific patient is found in only one fold. Data augmentation ensures that, within the same fold, there is the same number of data vectors for each class, i.e. the training data is perfectly balanced. The ResNet deep learning model achieved an overall accuracy of 97.25% for the six-class problem. For the medically significant arrhythmia no-arrhythmia problem, the same model achieved ACC=97.25%, SEN=89.18%, and SPE = 98.88%. The ROC curve shows that a 50% detection threshold yields the highest classification accuracy. Hence, the reported accuracy measures are optimal for the ResNet model.

Accuracy values above 96% indicate that it is possible to classify SVT, ST, SB, AFIB, AFL, and NSR RR interval signals. Our results were based on a diverse dataset with signals from 10093 subjects. The results are also robust, because they were achieved with data from unknown patients. As such these results are similar to performance measures from other arrhythmia detection methods, such as ECG analysis. However, the measurement effort for RR intervals is lower and the signals have a much-reduced data rate when compared with ECG signals. This leads to advantages when it comes to communicating, storing and processing the data. Hence, we are confident that RR interval signals can play a significant role when it comes to arrhythmia detection. We envision that the results of our study can serve as theoretical foundation for practical arrhythmia detection systems. These systems can take the form of an adjunct tool which supports both diagnosis and treatment monitoring. Having such a tool might lead to more and longer measurements which result in more arrhythmias being detected in the population. This could improve outcomes for patients through early diagnosis and adequate as well as ongoing treatment calibration.

Acronyms

ACC	Accuracy
A-V	Atrio-Ventricular
ADA	American Diabetes Association
AED	Automated External Defibrillator
AES	Advanced Encryption Standard
AFL	Atrial Flutter
AFIB	Atrial Fibrillation
AFDB	Atrial Fibrillation Database
ApEn	Approximate Entropy
ANS	Autonomic Nervous System
ANN	Artificial Neural Network
ANSI	American National Standards Institute
ANT	Anterior left atrial free wall
AHI	Apnea/Hypopnea Index
ALLSTAR	Allostatic State Mapping by Ambulatory ECG Repository
AI	Artificial Intelligence
ASM	Active Shape Model
ANOVA	Analysis Of Variance
API	Application Programming Interface
AR	Autoregressive
ARMA	Autoregressive Moving Average
AT	Atrial Tachycardia
ATM	Automated Teller Machine
AUC	Area Under Curve
AVN	AtrioVentricular Node
AVNRT	Atrioventricular Node Reentrant Tachycardia
AVRT	Atrioventricular Reentrant Tachycardia
BA	Bayesian Averaging
BCI	Brain Computer Interface
BLE	Bluetooth Low Energy
BMI	Body Mass Index
BPA	Back-Propagation Algorithm
BPSO	Binary Particle Swarm Optimization
BPM	Beats Per Minute
BSN	Biomedical Sensor Network
CAD	Computer-Aided-Diagnosis
CaD	Capacity Dimension
CAN	Cardiovascular Autonomic Neuropathy
CART	Classification And Regression Tree
CD	Correlation Dimension
CEUS	Contrast Enhanced Ultrasound
CI	Confidence Interval
CHF	Congestive Heart Failure
CHS	Community Health Centers
DL	Deep Learning
CLDA	Clustering Linear Discriminant analysis Algorithm
CNN	Convolutional Neural Network
CM	Clustered Microcalcifications
CSP	Communicating Sequential Processes
CS	Compressed Sampling
CT	Curvelet Transform
CPA	Communicating Process Architectures
CPC	Cardiopulmonary Coupling
CPU	Central Process Unit
CSA	Central Sleep Apnea
CSME	Clinically Significant Macular Edema
CSR	Cambridge Silicon Radio
CVD	Cardiovascular Disease
CWT	Continuous Wavelet Transform
D2H2	Distributed Diagnosis and Home Healthcare
DAGSVM	Directed Acyclic Graph Support Vector Machine
DB	Database
DBN	Deep Belief Network
DCT	Discrete Cosine Transform
DET	Determinism
DFT	Discrete Fourier Transform
DII	Diabetic Integrated Index
DL	Deep Learning
DLLE	Diagonal Line Lengths Entropy
DM	Diabetes Mellitus
DN	Diabetic Neuropathy
DNN	Deep Neural Network
DR	Diabetic Retinopathy
DT	Decision Tree
DWT	Discrete Wavelet Transform
E	External
E-M	Expectation–Maximization
ECG	Electrocardiogram
EDR	ECG-Derived Respiration
ECHONET	EchoCardiographic Healthcare Online Networking Expertise in Tasmania
EEG	Electroencephalogram
EM	Electromagnetic
EMD	Empirical Mode Decomposition
EMG	Electromyogram
EOG	Electrooculogram
FAQ	Frequently Asked Questions
FB	Fusion Beat
FD	Fractal Dimension
FDDI	Fiber Distributed Data Interface
FEn	Fuzzy Entropy
FES	Functional Electrical Stimulation
FCM	Fuzzy C-Means
FIR	Finite Impulse Response
FN	False Negative
FNN	False Nearest Neighbor
FP	False Positive
FSC	Fuzzy Sugeno Classifier
FT	Fourier Transform
GA	Genetic Algorithm
GLCM	Gray-Level Co-occurrence Matrix
GMM	Gaussian Mixture Model
GOE	Global Observatory for eHealth
GPRS	General Packet Radio Service
GPU	Graphics Processing Unit
GSM	Global System for Mobile communications
GUI	Graphical User Interface
H	Hurst exponent
HDNT	Hospital Digital Networking Technologies
HIHM	Home Integrated Health Monitor
HIPAA	Health Insurance Portability and Accountability Act
HMA	Haemorrhages And Microaneurysms
HMM	Hidden Markov Models
HOS	Higher Order Spectra
HR	Heart Rate
HRUS	High Resolution Ultrasound

(continued on next column)

(continued)

(continued on next page)

(*continued*)

HRV	Heart Rate Variability
HRVAS	Heart Rate Variability Analysis Software
HSDPA	High-Speed Downlink Packet Access
I	Internal
ICA	Independent Component Analysis
IEEE	Institute of Electrical and Electronic Engineers
IMF	Intrinsic Mode Functions
IoT	Internet of Things
IIR	Infinite Impulse Response
IR	Infrared
IT	Information Technology
IVUS	Intravascular Ultrasound
JMMB	Journal of Mechanics in Medicine and Biology
KS	Kolmogorov Sinai
K-NN	K-Nearest Neighbor
LA	Left Atrium
LAM	Laminarity
LBP	Local Binary Pattern
LC	Linear Classifier
LCP	Local Configuration Pattern
LDA	Linear Discriminant Analysis
LFP	Local Field Potentials
LIP	Left Inferior Pulmonary
LLE	Largest Lyapunov Exponent
LMNN	Levenberg–Marquardt Neural Network
LSP	Left Superior Pulmonary
LSTM	Long Short-Term Memory
LTAFDB	Long-Term AF Database
LTE	Long Term Evolution
LRNC	Lipid-Rich Necrotic Core
LV	Left Ventricular
MA	Moving Average
MC	Clustered Microcalcifications
NCSME	Non-Clinically Significant Macular Edema
MCT	Mobile Cardiac Telemetry
MDA	Multiple Discriminant Analysis
NDDF	Normal Density Discriminant Function
MI	Myocardial Infarction
ML	Machine Learning
MLPNN	Multilayer Perceptron Neural Networks
MMSE	Modified Multiscale Entropy
MOH	Ministry Of Health
MP	Mobile Phone
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MSE	Multiscale Entropy
NASA	National Aeronautics and Space Administration
NB	Naïve Bayes
nB	Number of Beats
NEB	Non-Ectopic Beat
nP	Number of Patients
NPDR	Non-Proliferative Diabetic Retinopathy
NSR	Normal Sinus Rhythm
OFDM	Orthogonal Frequency-Division Multiplexing
OSA	Obstructive Sleep Apnea
PAF	Paroxysmal Atrial Fibrillation
PCA	Principal Component Analysis
PDR	Proliferative Diabetic Retinopathy
PET	Positron Emission Tomography
PET-MRI	Positron Emission Tomography - Magnetic Resonance Imaging
PHC	Primary Health Centers
PNN	Probabilistic Neural Network
PPG	Photoplethysmogram
POS	Posterior left atrial free wall
PPV	Positive Predictive Value
PRE	Precision
PSD	Power Spectral Density
PSG	Polysomnography
PVC	Premature Ventricular Contraction
QoS	Quality of Service
RAM	Random Access Memory
RBF	Radial Basis Function
RBFNN	Radial Basis Function Neural Network
RBM	Restricted Boltzmann Machine
REC	Recurrence Rate
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
RF	Random Forest
RHN	Regional Health Network
RIP	Right Inferior Pulmonary
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RQA	Recurrence Quantification Analysis
RP	Recurrence Plot
RSA	Respiratory Sinus Arrhythmia
RSP	Right Superior Pulmonary
RVM	Relevance Vector Machine
S-D	Service-Dominant
SA	Sino-Atrial
SAAWR	Sinus Atrium to Atrial Wandering Rhythm
SAE	Sparse-Auto-Encoder
SampEn	Sample Entropy
SAN	SinoAtrial Node
SB	Sinus Bradycardia
SCD	Sudden Cardiac Death
SDNN	Standard Deviation of Normal to Normal intervals
SEN	Sensitivity
SE	Standard Error
ShanEn	Shannon Entropy
SI	Sinus Irregularity
SMDS	Switched Multimegabit Data Service
SNN	Spiking Neural Network
SOM	Self Organizing Map
SPE	Specificity
ST	Sinus Tachycardia
STFT	Short-Time Fourier Transform
SWT	Stationary Wavelet Transform
SQL	Structured Query Language
SVEB	Supraventricular Ectopic Beat
SVM	Support Vector Machine
SVT	Supraventricular Tachycardia
SWT	Stationary Wavelet Transform
T	Thoracic
TIA	Transient Ischemic Attack
TEE	Transesophageal Echocardiogram
TN	True Negative
TP	True Positive
TT	Trapping Time
TTE	Transthoracic Echocardiography
TVCF	Time-Varying Coherence Function
UB	Unclassified Beat
UIT	Urinary Tract Infection
ULP	Ultra-Low Power
UMTS	Universal Mobile Telecommunications System
US	Ultrasound
UWB	Ultra Wideband
VEB	Ventricular Ectopic Beat
VF	Ventricular Fibrillation
VFIB	Ventricular fibrillation
VFL	Ventricular Flutter
VT	Ventricular Tachycardia
WBAN	Wireless Body Area Network
WEP	Wired Equivalent Privacy
WHO	World Health Organization
WiMAX	Worldwide interoperability for Microwave Access
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WPA	Wi-Fi Protected Access
WPAN	Wireless Personal Area Network
WPD	Wavelet Packet Decomposition
WWAN	Wireless Wide Area Network

(*continued on next column*)

CRediT authorship contribution statement

Oliver Faust: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft. **U. Rajendra Acharya:** Writing - review & editing.

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.

References

Acharya, U.R., Fujita, H., Adam, M., Lih, O.S., Hong, T.J., Sudarshan, V.K., & Koh, J.E. (2016). Automated characterization of arrhythmias using nonlinear features from tachycardia ecg beats. 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 000533–000538.

Acharya, U. R., Fujita, H., Lih, O. S., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network. *Information sciences, 405*, 81–90.

Böhm, M., Frey, N., Giannitsis, E., Sliwa, K., & Zeiher, A. M. (2020). Coronavirus disease 2019 (covid-19) and its implications for cardiovascular care: Expert document from the german cardiac society and the world heart federation. *Clinical Research in Cardiology, 1*.

Camm, A. J., & Garratt, C. J. (1991). Adenosine and supraventricular tachycardia. *The New England Journal of Medicine, 325*(23), 1621–1629.

Castillo, O., Melin, P., Ramírez, E., & Soria, J. (2012). Hybrid intelligent system for cardiac arrhythmia classification with fuzzy k-nearest neighbors and neural networks combined with a fuzzy system. *Expert Systems With Applications, 39*(3), 2947–2955.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority oversampling technique. *Journal of artificial intelligence research, 16*, 321–357.

Chollet, F. (2015). Keras.

Chugh, S. S., Reinier, K., Teodorescu, C., Evanado, A., Kehr, E., Al Samara, M., Mariani, R., Gunson, K., & Jui, J. (2008). Epidemiology of sudden cardiac death: Clinical and research implications. *Progress in cardiovascular diseases, 51*(3), 213–228.

Clerkin, K. J., Fried, J. A., Raikhelkar, J., Sayer, G., Griffn, J. M., Masoumi, A., Jain, S. S., Burkhoff, D., Kumaraiah, D., & Rabbani, L. (2020). Covid-19 and cardiovascular disease. *Circulation, 141*(20), 1648–1655.

Clifford, G. D., Azuaje, F., & Mcsharry, P. (2006). Ecg statistics, noise, artifacts, and missing data. *Advanced methods and tools for ECG data analysis, 6*, 18.

Colilla, S., Crow, A., Petkun, W., Singer, D. E., Simon, T., & Liu, X. (2013). Estimates of current and future incidence and prevalence of atrial fibrillation in the us adult population. *The American journal of cardiology, 112*(8), 1142–1147.

Colucci, R. A., Silver, M. J., & Shubrook, J. (2010). Common types of supraventricular tachycardia: Diagnosis and management. *American family physician, 82*(8), 942–952.

De Caterina, R., Atar, D., Hohnloser, S. H., & Hindricks, G. (2012). 2012 focused update of the esc guidelines for the management of atrial fibrillation. *European Heart Journal, 33*, 2719–2747.

Delacrétaz, E. (2006). Supraventricular tachycardia. *The New England Journal of Medicine, 354*(10), 1039–1051.

Demski, A., & Soria, M. L. (2016). Ecg-kit: A matlab toolbox for cardiovascular signal processing. *Journal of open research software, 4*(1).

Desai, U., Martis, R. J., Acharya, U. R., Nayak, C. G., Seshikala, G., & Shetty, K. R (2016). Diagnosis of multiclass tachycardia beats using recurrence quantification analysis and ensemble classifiers. *Journal of Mechanics in Medicine and Biology, 16*(01), 1640005.

Driggin, E., Madhavan, M. V., Bikdeli, B., Chuich, T., Laracy, J., Biondi-Zoccai, G., Brown, T. S., Der Nigoghossian, C., Zidar, D. A., & Haythe, J. (2020). Cardiovascular considerations for patients, health care workers, and health systems during the covid-19 pandemic. *Journal of the American College of Cardiology, 75*(18), 2352–2371.

Drummond, C., & Holte, R. C. (2003). C4. 5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling. *Workshop on learning from imbalanced datasets II, 11*, 1–8.

Eraut, D., & Shaw, D. B. (1971). Sinus bradycardia. *British Heart Journal, 33*(5), 742.

Faust, O., Acharya, U. R., & Tamura, T. (2012). Formal design methods for reliable computer-aided diagnosis: A review. *IEEE reviews in biomedical engineering, 5*, 15–28.

Faust, O., Barika, R., Shenfield, A., Ciaccio, E.J., & Acharya, U.R. (2020). Accurate detection of sleep apnea with long short-term memory network based on rr interval signals. Knowledge-Based Systems, Ahead of print, 1–31.

Faust, O., Ciaccio, E. J., & Acharya, U. R. (2020). A review of atrial fibrillation detection methods as a service. *International Journal of Environmental Research and Public Health, 17*(9), 3093.

Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer methods and programs in biomedicine, 161*, 1–13.

Faust, O., Lei, N., Chew, E., Ciaccio, E. J., & Acharya, U. R. (2020). A smart service platform for cost effcient cardiac health monitoring. *International Journal of Environmental Research and Public Health, 17*(17), 6313.

Faust, O., Shenfield, A., Kareem, M., San, T. R., Fujita, H., & Acharya, U. R. (2018). Automated detection of atrial fibrillation using long short-term memory network with rr interval signals. *Computers in biology and medicine, 102*, 327–335.

Faust, O., Yu, W., & Acharya, U. R. (2015). The role of real-time in biomedical science: A meta-analysis on computational complexity, delay and speedup. *Computers in biology and medicine, 58*, 73–84.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters, 27*(8), 861–874.

Ferguson, J. D., & DiMarco, J. P. (2003). Contemporary management of paroxysmal supraventricular tachycardia. *Circulation, 107*(8), 1096–1099.

Fisher, A. C., Eleuteri, A., Groves, D., & Dewhurst, C. J. (2012). The ornstein–uhlenbeck third-order gaussian process (ougp) applied directly to the un-resampled heart rate variability (hrv) tachogram for detrending and low-pass filtering. *Medical & biological engineering & computing, 50*(7), 737–742.

Fox, D. J., Tischenko, A., Krahn, A. D., Skanes, A. C., Gula, L. J., Yee, R. K., & Klein, G. J. (2008). Supraventricular tachycardia: Diagnosis and management. *Mayo Clinic Proceedings, 83*(12), 1400–1411.

Friberg, L., Rosenqvist, M., Lindgren, A., Terént, A., Norrving, B., & Asplund, K. (2014). High prevalence of atrialfibrillation among patients with ischemic stroke. *Stroke, 45*(9), 2599–2605.

Fujita, H., & Cimr, D. (2019). Computer aided detection forfibrillations and utters using deep convolutional neural network. *Journal of Information Science, 486*, 231–239.

Furberg, C.D., Psaty, B.M., Manolio, T.A., Gardin, J.M., Smith, V.E., Rautaharju, P.M., & CHS Collaborative Research Group. (1994). Prevalence of atrialfibrillation in elderly subjects (the cardiovascular health study). The American journal of cardiology, 74(3), 236–241.

Fuster, V., Rydén, L. E., Asinger, R. W., Cannom, D. S., Crijns, H. J., Frye, R. L., Halperin, J. L., Kay, G. N., Klein, W. W., & Lévy, S. (2001). Acc/aha/esc guidelines for the management of patients with atrialfibrillation: Executive summary: A report of the american college of cardiology/american heart association task force on practice guidelines and the european society of cardiology committee for practice guidelines and policy conferences (committee to develop guidelines for the management of patients with atrialfibrillation) developed in collaboration with the north american society of pacing and electrophysiology. *Journal of the American College of Cardiology, 38*(4), 1231–1265.

Fuster, V., Rydén, L. E., Cannom, D. S., Crijns, H. J., Curtis, A. B., Ellenbogen, K. A., Halperin, J. L., Kay, G. N., Le Huezey, J.-Y., & Lowe, J. E. (2011). 2011 accf/aha/hrs focused updates incorporated into the acc/aha/esc 2006 guidelines for the management of patients with atrialfibrillation: A report of the american college of cardiology foundation/american heart association task force on practice guidelines. *Circulation, 123*(10), e269–e367.

Ganatra, S., Hammond, S.P., & Nohria, A. (2020). The novel coronavirus disease (covid-19) threat for patients with cardiovascular disease and cancer. Ganz, L.I., & Friedman, P.L. (1995). Supraventricular tachycardia. New England Journal of Medicine, 332 (3), 162–173.

Ganz, L. I., & Friedman, P. L. (1995). Supraventricular tachycardia. *New England Journal of Medicine, 332*(3), 162–173.

Ghali, W. A., Wasil, B. I., Brant, R., Exner, D. V., & Cornuz, J. (2005). Atrial utter and the risk of thromboembolism: A systematic review and meta-analysis. *The American journal of medicine, 118*(2), 101–107.

Gillis, A. M., Krahn, A. D., Skanes, A. C., & Nattel, S. (2013). Management of atrialfibrillation in the year 2033: New concepts, tools, and applications leading to personalized medicine. *The Canadian Journal of Cardiology, 29*(10), 1141–1146.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdor, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., & Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation, 101*(23), e215–e220.

Hamed, I., & Owis, M. I. (2016). Automatic arrhythmia detection using support vector machine based on discrete wavelet transform. *Journal of Medical Imaging and Health Informatics, 6*(1), 204–209.

Hanin, B. (2018). Which neural net architectures give rise to exploding and vanishing gradients? *Advances in Neural Information Processing Systems*, 582–591.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering, 21*(9), 1263–1284.

Henzel, N., Wróbel, J., & Horoba, K. (2017). Atrialfibrillation episodes detection based on classification of heart rate derived features. In *2017 MIXDES-24th International Conference Mixed Design of Integrated Circuits and Systems* (pp. 571–576).

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery, 33*(4), 917–963.

Ivanovic, M. D., Atanasoski, V., Shvilkin, A., Hadzievski, L., & Maluckov, A. (2019). Deep learning approach for highly specific atrialfibrillation and utter detection based on rr intervals. In *2019 41st Annual Inter-national Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1780–1783).

Iyengar, N., Peng, C., Morin, R., Goldberger, A. L., & Lipsitz, L. A. (1996). Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 271*(4), R1078–R1084.

Jensen, P. N., Gronroos, N. N., Chen, L. Y., Folsom, A. R., Defilippi, C., Heckbert, S. R., & Alonso, A. (2014). Incidence of and risk factors for sick sinus syndrome in the general population. *Journal of the American College of Cardiology, 64*(6), 531–538.

Jiang, A.H., Wong, D.L. -K., Zhou, G., Andersen, D.G., Dean, J., Ganger, G.R., Joshi, G., Kaminksy, M., Kozuch, M., Lipton, Z.C., & Pillai, P. (2019). Accelerating deep learning by focusing on the biggest losers.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data, 6*(1), 27.

Kochi, A. N., Tagliari, A. P., Forleo, G. B., Fassini, G. M., & Tondo, C. (2020). Cardiac and arrhythmic complications in patients with covid-19. *Journal of Cardiovascular Electrophysiology, 31*(5), 1003–1008.

Krijthe, B. P., Kunst, A., Benjamin, E. J., Lip, G. Y., Franco, O. H., Hofman, A., Witteman, J. C., Stricker, B. H., & Heeringa, J. (2013). Projections on the number of

individuals with atrial fibrillation in the european union, from 2000 to 2060. *European heart journal, 34*(35), 2746–2751.
Laguna, P., Moody, G. B., & Mark, R. G. (1998). Power spectral density of unevenly sampled data by leastsquare analysis: Performance and application to heart rate signals. *IEEE Transactions on Biomedical Engineering, 45*(6), 698–715.
Lainscak, M., Dagres, N., Filippatos, G. S., Anker, S. D., & Kremastinos, D. T. (2008). Atrial fibrillation in chronic non-cardiac disease: Where do we stand? *International journal of cardiology, 128*(3), 311–315.
Lashgari, E., Liang, D., & Maoz, U. (2020). Data augmentation for deep-learning-based electroencephalography. *Journal of Neuroscience Methods, 108885*.
Lei, N., Kareem, M., Moon, S. K., Ciaccio, E. J., Acharya, U. R., & Faust, O. (2021). Hybrid decision support to monitor atrial fibrillation for stroke prevention. *International Journal of Environmental Research and Public Health, 18*(2), 813.
Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Jia, Y., Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Xiaoqiang Zheng. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems [Software available from tensorow.org]. http://tensorow.org/.
Martínez, J. P., Almeida, R., Olmos, S., Rocha, A. P., & Laguna, P. (2004). A wavelet-based ecg delineator: Evaluation on standard databases. *IEEE Transactions on biomedical engineering, 51*(4), 570–581.
Melin, P., Miramontes, I., & Prado-Arechiga, G. (2018). A hybrid model based on modular neural networks and fuzzy systems for classification of blood pressure and hypertension risk diagnosis. *Expert Systems With Applications, 107*, 146–164.
Moody, G. B., & Mark, R. G. (1983). A new method for detecting atrialfibrillation using r-r intervals. *Computers in Cardiology, 10*(-), 227–230.
Moody, G. B., & Mark, R. G. (2001). The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine, 20*(3), 45–50.
Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., Das, S. R., De Ferranti, S., Després, J.-P., & Fullerton, H. J. (2016). Executive summary: Heart disease and stroke statistics 2016 update: A report from the american heart association. *Circulation, 133*(4), 447–454.
Muthuchudar, A., & Baboo, S. S. (2013). A study of the processes involved in ecg signal analysis. *International Journal of Scientific and Research Publications, 3*(3), 1–5.
Naccarelli, G. V., Varker, H., Lin, J., & Schulman, K. L. (2009). Increasing prevalence of atrial fibrillation and utter in the united states. *The American journal of cardiology, 104*(11), 1534–1539.
Olshansky, B., & Sullivan, R. M. (2013). Inappropriate sinus tachycardia. *Journal of the American College of Cardiology, 61*(8), 793–801.
Orejarena, L. A., Vidaillet, H., DeStefano, F., Nordstrom, D. L., Vierkant, R. A., Smith, P. N., & Hayes, J. J. (1998). Paroxysmal supraventricular tachycardia in the general population. *Journal of the American College of Cardiology, 31*(1), 150–157.
Petrėnas, A., Marozas, V., & Sörnmo, L. (2015). Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Computers in biology and medicine, 65*, 184–191.
Pudukotai Dinakarrao, S. M., & Jantsch, A. (2018). Addhard: Arrhythmia detection with digital hardware by learning ecg signal. In *Proceedings of the 2018 on Great Lakes Symposium on VLSI* (pp. 495–498).
Ramirez, E., Melin, P., & Prado-Arechiga, G. (2019). Hybrid model based on neural networks, type-1 and type-2 fuzzy systems for 2-lead cardiac arrhythmia classification. *Expert Systems With Applications, 126*, 295–307.
Salem, M., Taheri, S., & Yuan, J. (2018). Ecg arrhythmia classification using transfer learning from 2- dimensional deep cnn features. *IEEE Biomedical Circuits and Systems Conference (BioCAS), 2018*, 1–4.
Sanoski, C.A. (2009). Clinical, economic, and quality of life impact of atrial fibrillation. Journal of Managed Care Pharmacy, 15 (6 Supp B), 4–9.
Sawhney, N. S., & Feld, G. K. (2008). Diagnosis and management of typical atrial utter. *Medical Clinics of North America, 92*(1), 65–85.
Scholar, E. (2015). Xpharm: The comprehensive pharmacology reference. Elsevier Inc, 10, B978–008055232.
Semelka, M., Gera, J., & Usman, S. (2013). Sick sinus syndrome: A review. *American family physician, 87*(10), 691–696.
Still, A.-M., Raatikainen, P., Ylitalo, A., Kauma, H., Ikäheimo, M., Kesäniemi, Y. A., & Huikuri, H. V. (2005). Prevalence, characteristics and natural course of inappropriate sinus tachycardia. *EP Europace, 7*(2), 104–112.
Taha, B., Reddy, S., Xue, Q., & Swiryn, S. (2000). Automated discrimination between atrial fibrillation and atrial utter in the resting 12-lead electrocardiogram. *Journal of electrocardiology, 33*, 123–125.
Tanne, J. H., Hayasaki, E., Zastrow, M., Pulla, P., Smith, P., & Rada, A. G. (2020). Covid-19: How doctors and healthcare systems are tackling coronavirus worldwide. *Bmj, 368*.
Wolf, P. A., Abbott, R. D., & Kannel, W. B. (1991). Atrialfibrillation as an independent risk factor for stroke: The framingham study. *Stroke, 22*(8), 983–988.
Xia, Y., Wulan, N., Wang, K., & Zhang, H. (2018). Detecting atrialfibrillation by deep convolutional neural networks. *Computers in biology and medicine, 93*, 84–92.
Yildirim, O., Talo, M., Ciaccio, E. J., San Tan, R., & Acharya, U. R. (2020). Accurate deep neural network model to detect cardiac arrhythmia on more than 10,000 individual subject ecg records. In *Computer methods and programs in biomedicine* (p. 105740).
Yuan, C., Yan, Y., Zhou, L., Bai, J., & Wang, L. (2016). Automated atrialfibrillation detection based on deep learning network. *IEEE International Conference on Information and Automation (ICIA), 2016*, 1159–1164.
Zheng, J., Zhang, J., Danioko, S., Yao, H., Guo, H., & Rakovski, C. (2020). A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. *Scientific Data, 7*(1), 1–8.
Zhou, X., Ding, H., Ung, B., Pickwell-MacPherson, E., & Zhang, Y. (2014). Automatic online detection of atrialfibrillation based on symbolic dynamics and shannon entropy. *Biomedical engineering online, 13*(1), 18.