Expert Systems With Applications 203 (2022) 117206

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Review

Comprehensive survey of computational ECG analysis: Databases, methods and applications

Elena Merdjanovska [a,b], Aleksandra Rashkovska [a,*]

[a] Department of Communication Systems, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
[b] Jožef Stefan International Postgraduate School, Jamova 39, 1000 Ljubljana, Slovenia

A R T I C L E I N F O

Keywords:
ECG
Mobile
Signal processing
Database
Machine learning
Survey

A B S T R A C T

Electrocardiogram (ECG) recordings are indicative for the state of the human heart. Automatic analysis of these recordings can be performed using various computational methods from the areas of signal processing and machine learning. In addition to the 12-lead ECG devices and the Holter monitor, as currently the most widely used ECG screening methods in clinical practice, ECG recordings are recently often acquired with small novel wireless ECG body sensors. These novel types of body sensors allow for ECG monitoring and analysis to be used for a much broader array of applications than only diagnosing cardiovascular disorders. The new types of ECG measuring devices, as well as their broader and more frequent use, pose new challenges in the processing and analysis of ECG, and furthermore, raise the need for automatic, low-cost, real-time, and efficient ECG monitoring that can be used at home or under ambulatory settings alike. This paper provides a comprehensive survey on the variety of both ECG data and computational methods in various applications: morphological and rhythmic arrhythmia detection, signal quality assessment, biometric identification, respiration estimation, fetal ECG extraction, and physical and emotional monitoring. It includes an extensive overview of 45 diverse ECG public databases and their analysis with state-of-the-art computational ECG methods. We highlight the most notable achievements in each of these ECG application areas in the recent years, and, furthermore, identify future trends in computational ECG analysis, especially analysis of ECG from mobile devices. The general conclusion is that ECG for medical diagnosis is successfully analyzed with the existing methods, while different applications during daily ECG monitoring are still open fields. Given how deep learning has been able to successfully address a lot of the most significant computational ECG problems, like arrhythmia classification, in future, it is expected for deep learning methods to be comprehensively tested in areas where they have not been yet applied, such as respiration estimation and fetal ECG extraction.

1. Introduction

Electrocardiography is the most common and extensively used vital sign monitoring process in modern healthcare systems. Electrocardiogram (ECG) recordings capture the electric potential on the body surface, which is a result of propagation of the electrical signal in the heart (Trobec et al., 2018). Consequently, many cardiac abnormalities have a signature on the ECG signal and their identification can help diagnose cardiac disorders. In 2019, cardiovascular disorders caused 32% of all deaths in the world according to the World Health Organization (WHO, 2021), which makes them a major burden worldwide. Therefore, early detection of the patients at risk, monitoring of the diagnosed patients, and a better understanding of the disease mechanisms are crucial for improving diagnosis and treatment.

ECG can be recorded in different formats: standard 12-lead ECGs provide information on cardiac activity from 12 different perspectives (leads) during a short time, whereas Holter ECGs record the electrical activity of the heart over longer periods of time (several hours) from 5–7 leads. In addition to these two methods, which are currently the

* Corresponding author.
E-mail addresses: elena.merdjanovska@ijs.si (E. Merdjanovska), aleksandra.rashkovska@ijs.si (A. Rashkovska).

https://doi.org/10.1016/j.eswa.2022.117206
Received 24 September 2020; Received in revised form 1 December 2021; Accepted 7 April 2022
Available online 23 April 2022

Fig. 1. Savvy ECG sensor in use.

widely used ones in clinical practice, small novel wireless ECG body sensors are being developed. The market of medical-grade wearable ECG devices has expended during the last years, including products from the big players in the electronic industry, like Samsung with S-Patch Cardio[1] and Philips with their wearable biosensor,[2] and even more, products by companies specialized in medical equipment, like ZIO by iRhythm,[3] VitalPatch by VitalConnect,[4] KardiaMobile ECG by Qardia,[5] and Savvy ECG by Saving.[6] As an example, the Savvy ECG sensor, developed at the Jožef Stefan Institute (Rashkovska et al., 2020), is presented in Fig. 1. These sensors are intended for long-term monitoring and provide single-lead ECG by measuring the electrical potential difference between proximal electrodes placed on the skin near the heart, accordingly also called differential lead. These types of body sensors allow for ECG analysis and monitoring to be used for a much broader array of applications than only diagnosing cardiovascular disorders. The provision of mobile health (mHealth) services, like patient monitoring in hospitals, remote medical support, or monitoring during sport activities are some of the newly established areas. In addition, ECG body sensors usually combine different functions into in a single physical device, and are consequently often referred to as multi-functional body sensors (Trobec et al., 2014).

The new types of ECG measuring devices, as well as their broader and more frequent use, pose new challenges. Some of them are related to hardware and communications aspects, such as the need to allow greater patient mobility and provide wireless transmission of the data from the device to a nearby personal terminal (Rashkovska et al., 2020). Other challenges arise in the processing and analysis of the signals. Namely, large amounts of ECG data are being recorded using these new measurement devices and manually studying such large amounts of ECG data can be tedious and time-consuming. This increases the need for automatic, low-cost, real-time, and efficient ECG monitoring that can be used at home or under ambulatory settings alike. Therefore, there is a need for powerful computational methods to maximize the information extracted from comprehensive ECG data. The variety of ECG data and their applications also calls for a diversity of algorithms to address this need (Lyon et al., 2018).

This paper provides a comprehensive survey of the variety of both ECG data and algorithms. First, it sets the ground base by providing an overview of the processing methods and algorithms that form the ECG analysis pipeline: denoising, segmentation, feature extraction and selection, learning algorithms and evaluation methods. After that, the paper mainly focuses on an overview of the experimental setups and results for a broad range of ECG application tasks: morphological and rhythmic arrhythmia detection, signal quality assessment, biometric identification, respiration estimation, fetal ECG extraction, and physical and emotional monitoring. It also includes an overview of commercial ECG analysis software. Furthermore, a critical review of the methods and data used in the encompassed research is also provided, concluding with final remarks and future challenges in computational ECG analysis.

Related survey studies on ECG analysis mostly focus on one application area, most often arrhythmia detection (Dinakarrao et al., 2019; Hoefman et al., 2010; Luz et al., 2015), and lately also on respiration estimation from ECG (Charlton et al., 2018). Other application areas have been surveyed very little or not at all. Some survey studies focus only on one methodology, in particular, the very popular deep learning approach (Faust et al., 2018). Compared to previous survey studies, the main significance of this paper is recognized in the variety of public ECG databases covered as well as the opportunities for their comprehensive analysis in light of different application areas as end-tasks. Moreover, the focus of this paper is on the newest trends in the ECG analysis field, mainly ECG from mobile devices, attempting to identify the directions in which ECG analysis is headed.

The main contributions of this paper can be summarized as follows:

- Survey of computational methods used in different stages of the ECG analysis pipeline: denoising, segmentation, feature extraction and selection, learning algorithms, and evaluation methods,
- Extensive overview of 45 most significant and popular databases containing various ECG recordings,
- Highlighting the most notable achievements in recent years for 7 ECG application areas: morphological and rhythmic arrhythmia detection, signal quality assessment, biometric identification, respiration estimation, fetal ECG extraction, and physical and emotional monitoring,
- Identifying future trends in computational ECG analysis, especially analysis of ECG from mobile devices.

[1] https://www.wellysis.com/.
[2] https://www.usa.philips.com/healthcare/product/HC989803196871/wearable-biosensor-wireless-remote-sensing-device.
[3] https://www.irhythmtech.com/.
[4] https://vitalconnect.com/.
[5] https://qardia.com.au/for-patients/kardiamobile-ecg-services.
[6] http://savvy.si/.

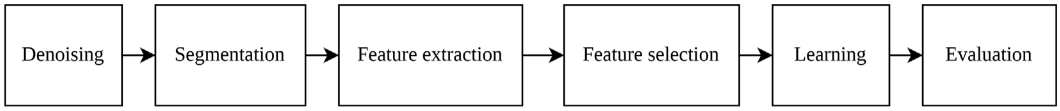

Fig. 2. ECG analysis pipeline.

The main goal of the critical overview is to research which kinds of ECG data are already well-researched and which data-collection scenarios are still to be explored.

2. ECG analysis pipeline

This section presents the algorithms used in different stages of the ECG analysis pipeline presented in Fig. 2. The pipeline includes the following steps: preprocessing (including denosing and segmentation), feature extraction and selection, learning algorithms, and finally, evaluation methods. It is important to note that not all of the steps are applicable in all application scenarios — this depends on the final task and the experimental setting.

2.1. Denoising

One of the first usual steps in the pipeline is reducing the noise in the ECG signal. Low frequency noise relates to signal baseline oscillations resulting from body movements and respiration, while power line interference and digitization of analog electrical potential result in high-frequency noise. In general, the most widely used approach to reduce the noise in ECG is the application of a standard Finite Impulse Response (FIR) filter to the signal, as performed in Raj and Ray (2018) and Trobec et al. (2012), or an Infinite Impulse Response (IIR) filter, most commonly Butterworth filters found in Da Poian et al. (2015). Usually, band-pass filters with varying cut-off frequencies are being used (with the low frequency starting as low as 0.1 Hz and the high one up to 100 Hz) (Luz et al., 2015). Other alternative approaches include Discrete Wavelet Transform (DWT) (Dinakarrao et al., 2019), Empirical Mode Decomposition (EMD) (Dinakarrao et al., 2019; Satija et al., 2019; Weng et al., 2006), which partitions the signal into multiple Intrinsic Mode Functions (IMFs), as well as patch-based methods such as the Non-Local Means (NLM) algorithm (Tracey & Miller, 2012). Furthermore, Bayesian filtering methods have also been explored for ECG denoising, more specifically methods aimed for nonlinear systems (Luz et al., 2015; Sameni et al., 2007), such as extended Kalman filters (Sayadi, 2008).

Adaptive techniques for ECG denoising are also well-researched. Such techniques are capable of tracking the signal under non-stationary conditions and accordingly adjusting the parameters of a denoising filter. An example thereof is the adaptive filter proposed in An and Stylios (2020), which shows better performance in motion reduction than other methods. Another example of an adaptive filter is the recent work on low-distortion adaptive Savitzky–Golay (LDASG) filtering method (Huang et al., 2019). This is a novel method based on a popular denoising technique used for the so-called Savitzky–Golay filtering, which is a FIR filter designed using polynomial approximation. The LDASG filtering method is discussed in more details in Section 3.3.

2.2. Segmentation

An important part of the pipeline is segmentation of the ECG signal. A large portion of the studies perform the end task on a beat-by-beat level, which means they look at each heartbeat separately during the analysis. Another option is to employ a fixed sliding window technique on the entire signal instead (Charlton, Bonnici et al., 2016; Hannun et al., 2019; Mousavi et al., 2019; Yao et al., 2020), with possibly a different number of beats in each window, or extract ECG sequences containing a specific number of successive beats (Šprager et al., 2017).

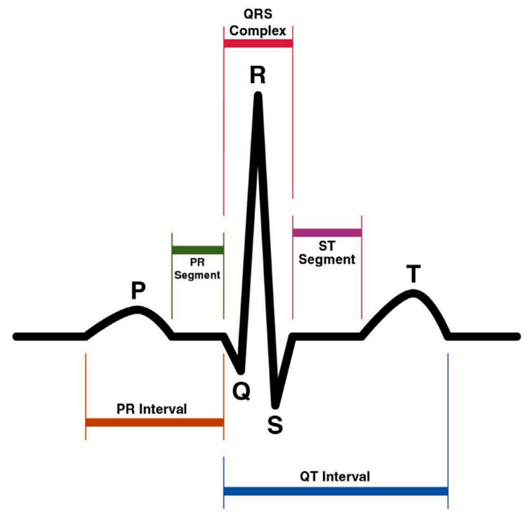

Fig. 3. Characteristic ECG heartbeat shape.

In this section, we will focus on methods that operate on beat-by-beat level and on QRS detection algorithms, which are crucial for determining the location of each beat.

Each heartbeat in an ECG signal is characterized with a specific waveform. This can be observed on Fig. 3, where we show an ECG heartbeat with its characteristic fiducial points (P, Q, R, S and T), and segments and duration intervals (PR interval, PR segment, QT interval, ST segment), as well as the QRS complex. Each of these entities corresponds to a specific phase in the cardiac cycle (Barrett et al., 2010), with the QRS complex corresponding to the most significant cardiac activity: the depolarization of the ventricles. This process is reflected in the ECG waveform with the largest potential differences, forming the R-peak. The location of an R-peak is highly important because it is used as an equivalent for the location of a heartbeat. Therefore, determining the R-peak locations is very significant in computational ECG analysis, mainly as a preprocessing step in beat-by-beat level classification algorithms (Luz et al., 2015), but also to perform additional analyses such as heart rate variability (HRV) analysis (Oweis & Al-Tabbaa, 2014).

The literature on detection of the most important ECG characteristic, i.e. the QRS complex, is vast, with many different approaches having been adopted in the last 30 years. One of the first QRS detectors with satisfactory performance is the one from Pan and Tompkins (Pan & Tompkins, 1985), based on analyses of slope, amplitude, and width of ECG waves. The Pan–Tompkins algorithm includes both low and high pass filtering for noise reduction, differentiation, squaring and moving window integration of the signal, ending by applying multiple adaptive thresholds to the peaks detected from the resulting waveform. Different methods based on the Pan–Tompkins algorithm have been developed subsequently, such as the popular Hamilton–Tompkins detector (Hamilton & Tompkins, 1986), as well as the Hamilton variant (Hamilton, 2002), with sensitivity and positive predictive value of up to 0.998.

In addition to improving performance, in recent years efforts have been made in developing faster QRS detection algorithms. Pan -Tompkins demonstrates high detection accuracy, but it has been shown that improvements regarding efficiency can be made by proposing simple-fast algorithms such as the optimized knowledge-based method in Elgendi (2013). They suggest a basic approach consisting of two moving averages calibrated by a knowledge base of ECG recordings, using only two parameters. Their method shows comparable detection results to Pan–Tompkins and its variants.

Lastly, it is worth mentioning that there have been some attempts to treat QRS detection as a separate machine learning task and use neural networks for localization of QRS complexes, more specifically multilayer perceptron (MLP) (Chromik et al., 2021) and convolutional neural networks (CNN) (Xiang et al., 2018) architectures. However, these cases are not representative of the standard ECG analysis pipeline, which this paper attempts to identify, since generally simple and fast QRS detectors are preferred as a first step in a pipeline for more complex ECG analysis tasks, most commonly arrhythmia detection.

2.3. Feature extraction and selection

There is a wide range of features that can be extracted from an ECG signal, in order to perform different end-tasks. When making a diagnosis, medical experts take into account the relative positions of the ECG components and their amplitudes, most notable of which is the time between the R peaks of consecutive heartbeats — the RR interval (Arzeno et al., 2008). RR intervals are necessary for HRV analysis, which is sometimes used for rhythm classification (Hirsch et al., 2021), but also in different tasks related to physical activity and emotion (Seshadri et al., 2019). Other features of this kind are the PR interval, the width of the QRS complex and the QT interval (Dinakarrao et al., 2019; Luz et al., 2015). The amplitude and duration of these shapes can sometimes suggest different problems with the heart, which is why the correct detection of these entities is of high importance. For example, the diagnosis of atrial arrhythmia depends on the presence or absence of P-waves, as well as their duration and timing (Macfarlane and et al., 2010). This group of features, which represents different statistical values derived from the fiducial ECG points, is referred to as temporal morphological features. Different methods, referred to as ECG segmentation algorithms, can be used to find these fiducial points, which generally incorporate some kind of QRS detection as their first step (Beraza & Romero, 2017), covered separately in the previous section.

Another approach is the extracting of features in the frequency domain by using Fast Fourier Transform (FFT) in order to discover changes in the power spectrum of the ECG waves (Dinakarrao et al., 2019). However, pure frequency-domain-based analysis is not so often used and more popular are methods operating in the time–frequency domain. Examples thereof are the application of Stockwell transform (Raj & Ray, 2018) and wavelet transform (Li et al., 2017; Qin et al., 2017) to ECG signals, whereby the coefficients obtained can be regarded as features of the ECG. More advanced statistical approaches can also be used for feature extraction from ECG, such as Higher Order Statistics (HOS) (Šprager et al., 2017), Autocorrelation-Linear Discriminant Analysis (AC/LDA) (Agrafioti & Hatzinakos, 2010) and Independent Component Analysis (ICA) (Da Poian et al., 2015). Some of these methods, mostly the transforms mentioned above, aim to model the ECG beat instead of directly extracting specific features.

As most of the methods discussed previously generate a large number of features, feature selection and dimensionality reduction are usually performed. This can be done with standard algorithms for feature ranking used in machine learning, such as the maximum Relevance Minimum Redundancy (mRMR) algorithm (Mei et al., 2018) and Sequential Feature Selection (SFS), as well as Principal Component Analysis (PCA) (Qin et al., 2017) and Singular Value Decomposition (SVD) (Šprager et al., 2017) for dimensionality reduction. In addition, in recent years, different stochastic optimization algorithms are being increasingly used for feature selection, e.g. Genetic Algorithm (GA) in Li et al. (2017) and Artificial Bee Colony (ABC) in Raj and Ray (2018).

2.4. Learning algorithms

This section presents the different learning algorithms and approaches commonly used in ECG analysis tasks. The methods covered vary from simple and explainable, such as rules, to the very complex deep neural networks. Furthermore, the section focuses on learning algorithms for supervised tasks, more specifically classification, as this is the most prominent type of end-task in recent studies.

The simplest type of diagnostic classification is using a set of rules. Developing rules usually requires high level of expert knowledge (strong involvement of medical professionals) since expert features are required. In addition, the rules or criteria obtained with this kind of analysis are explainable, which makes them desirable in clinical settings. Such an example is presented in Macfarlane et al. (2004), where a criterion for acute myocardial infarction is established from statistical analysis of ECG data and since then it has been adopted as standard in diagnostic practice. Furthermore, a successful rule-based beat classifier has been implemented by Hamilton (2002), which distinguishes normal from ventricular beats with a multi-stage algorithm utilizing beat matching techniques and beat morphology rules.

In addition to the simpler types of methods using rules and template matching, in the final classification step of the ECG analysis pipeline, a lot of different standard machine learning algorithms for classification can be used. There have been experiments with almost all existing classifier types in the past, however, most common and highest-performing in recent literature are Support Vector Machines (SVM) (Qin et al., 2017; Raj & Ray, 2018), with occasional use of Random Forest (Lyon et al., 2018), Nearest Neighbors Classifier (Šprager et al., 2017) and MLP networks (Chromik et al., 2021; Li et al., 2017). Furthermore, stochastic optimization methods are utilized in the parameter setting and tuning of the machine learning algorithms, which is an important step towards achieving best possible performance in the end-task. Examples thereof are the use of ABC in Raj and Ray (2018) and GA in Li et al. (2017).

Deep neural networks are probably the most researched learning algorithms right now — they are also very prominent in the area of ECG analysis. For all types of signals and images, including ECG, the architectures used are built with convolutional layers in their basis and are, therefore, called convolutional neural networks (CNN). In addition to all deep networks being some type of a CNN (Kiranyaz et al., 2017; Smith, Walsh et al., 2019), different additional mechanisms have been employed. A common example thereof are the so-called residual neural networks (ResNet), built by introducing residual blocks (Hannun et al., 2019; Ribeiro et al., 2020) consisting of multiple convolutional layers, with additional skip-connections which jump over blocks. Skipping over layers speeds up the training by reducing the problem of vanishing gradients. Other mechanism make use of the temporal dependency of both the ECG signal and its corresponding classes, such as Long Short-Term Memory (LSTM) (Chauhan & Vig, 2015; Yao et al., 2020), Gated Recurrent Units (GRU) (Chen et al., 2020) and Attention mechanisms (Chen et al., 2020; Hong et al., 2019; Mousavi et al., 2019; Yao et al., 2020).

An example of a complex deep learning architecture, combining a few of the mentioned advanced deep learning mechanisms, can be observed in Fig. 4. The architecture shown on this figure is taken from the study in Mousavi et al. (2019) and it is used for classification, more specifically arrhythmia detection. It includes both attention mechanism and LSTM units. Encoder–decoder architectures are also very popular for some ECG tasks, like heartbeat classification for arrhythmia detection (Mousavi & Afghah, 2019) and fetal ECG signal denoising (Fotiadou et al., 2020). They are especially suitable for sequence-to-sequence mapping when a signal as time-series is required at the output. On another hand, while deep learning methods are mainly used as end-tasks, they are also very successful as feature extractors combined with simple classifiers such as 1-Nearest Neighbor, as has been shown in Labati et al. (2018).

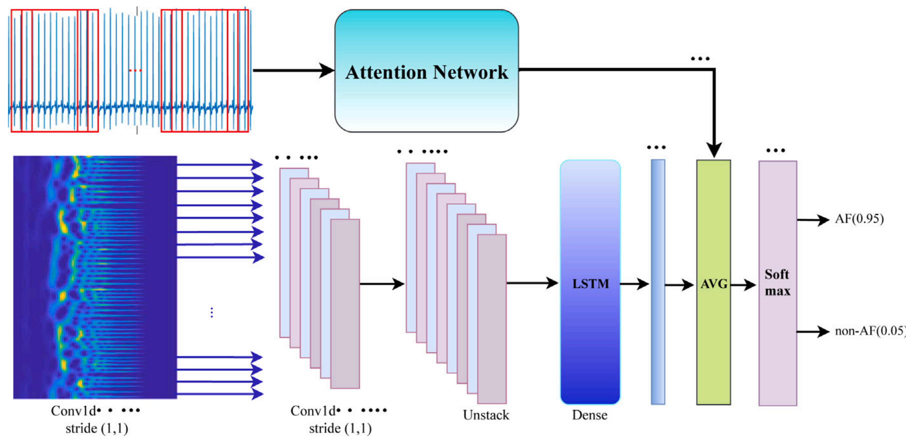

Fig. 4. Example deep learning architecture, combining CNN, LSTM and attention (Mousavi et al., 2019).

2.5. Evaluation methods

Different evaluation schemes can be used to evaluate supervised learning tasks for ECG analysis. The most important distinction when designing an evaluation setup for any kind of medical application is whether data from the same patients, which form the training dataset, is present in the testing dataset or not. Having this in mind, the Association for Advancement of Medical Instruments (AAMI) has proposed guidelines for evaluation of arrhythmia detection methods in the ANSI/AAMI EC57 standard (ANSI/AAMI EC57, 2012). Similar recommendations are also part of the IEC 60601-2-47 standard (IEC 60601-2-47, 2012). While the AAMI EC57 standard is used in FDA clearance, the IEC 60601-2-47 standard is widely used in CE mark certification. According to the standards, evaluation can be performed using the intra-patient (Mousavi & Afghah, 2019; Qin et al., 2017), inter-patient (Mousavi & Afghah, 2019; Qin et al., 2017; Raj & Ray, 2018; Zhai et al., 2020) or patient-specific (Kiranyaz et al., 2017; Raj & Ray, 2018; Zhai et al., 2020) schemes. The inter-patient paradigm reflects most real-life use cases, where the algorithms need to perform well when the analysis platform is utilized for entirely new people, while the patient-specific one allows for personalized approaches in ECG analysis. This general idea is applicable to almost all supervised tasks in the domain of medicine, however, these paradigms are originally established for arrhythmia detection using the most utilized database — MIT-BIH Arrhythmia (Moody & Mark, 2001). In the case of larger databases, mainly those that do not contain beat-by-beat annotations, such as the PTB-XL database (Wagner et al., 2020), the evaluation is exclusively inter-patient since a large enough number of distinct patients is present and other paradigms bring no benefits.

The inter-patient paradigm used in most studies splits the MIT-BIH dataset into two sets called DS1 and DS2. This split has been proposed in de Chazal et al. (2004) and has been adopted in practice as a standard extension of the inter-patient evaluation paradigm. DS1 (consisting of recordings from patients with IDs: 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223 and 230) is used to train classification models and DS2 (containing patient IDs: 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233 and 234) is used only for testing. For the patient-specific paradigm, the dataset is again split into 2 groups, the first one containing recordings from patients with IDs starting with 1 (100, 101, 103...), while the second one contains the recordings from patient with IDs starting with 2 (200, 201, 202...). The first group is used to form a set of common train heartbeats, while a separate training dataset is built for each patient in the second group. These patient-specific datasets consist of the common train heartbeats, together with beats from the first 5 min from the specific patient's recording. The remaining 25 min of each patient's recording are used to test the classification models. The third standardized evaluation paradigm is intra-patient, which includes a random split of all heartbeats into train and test, disregarding which patient they come from.

Annotating ECG data is a relatively expensive process, however, collecting ECG data is fairly simple. For this reason, semi-supervised learning is very suitable for ECG analysis because it allows to include unlabeled recordings in the training phase and it has been shown that this could enhance classification performance. Semi-supervised pipelines for ECG classification are most commonly performed as an improvement over the patient-specific evaluation, without the need for labels in the 5-min data from the specific patient. More specifically, the study in Zhai et al. (2020) uses a common label pool as well as iterative label estimation and label update over the unlabeled patient-specific data.

Another evaluation method used commonly in supervised learning, especially in limited-data situations such as ECG analysis, is cross-validation (Qin et al., 2017). Cross-validation can be used in different evaluation settings, even though, when it comes to the evaluation schemes for the MIT-BIH Arrhythmia database described above, random cross-validation is usually performed in the intra-patient schema, while inter-patient and patient-specific have pre-defined singular train-test splits. In addition, cross-validation can be performed as a part of parameter setting and initial evaluation of the models before performing testing on a separate test set, which is the case in the study in Raj and Ray (2018) where a 14-fold cross validation for parameter setting on the training dataset is executed. Related to limited number of data samples, and more importantly, imbalance in the number of samples of each class, different data augmentation techniques could be employed. An example thereof is the use of the SMOTE technique to increase the number of samples from less-represented classes in the training dataset (Mousavi & Afghah, 2019).

In addition to evaluation strategies, also important are the performance measures used to estimate and compare the performance of the methods. Classification tasks are usually evaluated using predictive accuracy, which is the portion of total samples correctly classified. Accuracy is however not the most reliable metric in the case of imbalanced problems, as are most medical applications, including ECG classification. Therefore, other measures, such as precision, recall, F1-score (Bramer, 2007), and area under the ROC curve (AUC) (Fawcett, 2006), are used as well. These metrics are defined for binary classification, or for each class separately in the case of multi-class classification. Since a large portion of the studies dealing with heartbeat classification for arrhythmia detection include more than two classes, per-class metrics need to be averaged into a single performance measure. Averaging can be performed in a macro, micro or weighted manner, with macro measures calculating a simple mean over all classes, and are considered most suitable for evaluating imbalanced multi-class classification cases (Strodthoff et al., 2021).

In application areas where heartbeat location is important, such as QRS detection and fetal ECG extraction, precision and recall are widely used to evaluate performance (ANSI/AAMI EC57, 2012). In addition, for fetal ECG extraction, the ANSI/AAMI EC57 standard allows for a ±50 ms acceptance interval between the detected and the closest reference annotation. Some QRS detection methods also consider the average time error between the detected and reference R peak (Arzeno et al., 2008).

3. Applications

This section will cover the most significant areas of application of ECG analysis present in the research field in most recent years. We will provide an overview of the recent advances in ECG analysis, focusing on the different data-collection scenarios and end-tasks. The studies include a range from already well-established commercial products to implementations still in the research phase. The areas will be examined in the following order: arrhythmia detection and rhythm abnormalities, signal quality assessment, biometric identification, respiration estimation, fetal ECG extraction, physical and emotional monitoring, and finally, commercial ECG analysis software.

Regarding the data used to perform the ECG analysis, there are a lot of publicly available databases containing ECG recordings. They are of high importance and studies analyzing data available to everyone have higher value because this allows for fair comparison. In Table 1, the most significant and popular databases have been summarized, alongside with information about: the type of the recordings they include, like sampling frequency, number of leads, length and size (in terms of number of recordings); annotations included (in terms of targeted application areas); and notable studies using them. We can see that there are different settings for recording ECG signals, ranging from the standard 12-lead resting ECG to long-term monitoring with the Holter monitor. In the table, the databases with higher recording frequency and 12 or 15 leads are obtained with the standard resting ECG monitor, while the ones with a few leads are obtained with the Holter monitor. Further, we can observe that the recordings are of various duration, ranging from 10 s up to 24 h, and different sampling frequencies, usually 250 Hz or higher. Most of the databases listed in the table are available through the PhysioNet repository,[7] which is an online platform for sharing medical data. Additionally, a few of the datasets can be found on other public repositories such as figshare,[8] Zenodo,[9] and IEEE Data Port.[10] The goal of this section is to provide an overview of these databases and put them into context of specific end-tasks (application areas) in the next subsections. The application end-tasks for which these datasets can be used are determined by the present annotation types, and the additional signals and information they contain besides the ECG.

3.1. Heartbeat classification for arrhythmia detection

Various heartbeat abnormalities are known as arrhythmias under one name. These abnormalities are detected by medical professionals using ECG due to its simplicity and non-invasive nature. The development of automatic ECG-based heartbeat classification and arrhythmia detection methods represents a large portion of the research involving computational methods for ECG analysis. There are two main categories of arrhythmias. The first type are called morphological arrhythmias and are characterized by the irregularity of a single heartbeat. The second type are the so-called rhythmic arrhythmias, characterized by a set of irregular heartbeats. This section focuses on automatic detection of morphological arrhythmia types, while the next one handles separately the problem of detecting the different types of ECG rhythms.

The most popular publicly available arrhythmia database is the MIT-BIH Arrhythmia database (Moody & Mark, 2001). It includes Holter recordings from 48 subjects, 23 healthy and 25 selected to include clinically significant arrhythmias. This database contains annotations for a large number of both rhythmic and morphological arrhythmias, however, it is mostly used to differentiate single-heartbeat (morphological) irregularities. The annotations are on the locations of the R-peaks of each heartbeat, which is usually the case for all PhysioNet datasets. In addition to MIT-BIH Arrhythmia, other databases, such as the American Heart Association (AHA) database[11] and the St. Petersburg Institute of Cardiological Technics 12-lead Arrhythmia (INCART) database (Tihonenko et al., 2007), are used in a lot of studies for this task. Furthermore, there are a few novel datasets that contain heartbeat form labels, in addition to other diagnostic statements, but they have not been used as extensively as the MIT-BIH Arrhythmia Detection database in morphological arrhythmia detection studies yet. Examples thereof are the PTB-XL database (Wagner et al., 2020) and the Shaoxing People's Hospital's 10,000 patients arrhythmia database (Zheng et al., 2020). Unlike the standard MIT-BIH datasets that contain beat-level annotations, this new generation of arrhythmia datasets provides a set of labels for each ECG recording. The recordings in these datasets are much shorter, usually around 10 s, and only the information that an arrhythmia is present is given. The exact locations of the irregular heartbeats are not given, which requires a modified approach since heartbeat segmentation methods are not necessary.

AAMI in the ANSI/AAMI EC57 standard provides guidelines for grouping the large number of arrhythmia types into 5 main classes (N - normal, S - supraventricular arrhythmia, V - ventricular arrhythmia, F - fusion beats and Q - unknown) (ANSI/AAMI EC57, 2012), which are employed in a large portion of the research (Chauhan & Vig, 2015; Kiranyaz et al., 2017; Mousavi & Afghah, 2019). Moreover, because the most significant arrhythmia classes are supraventricular ectopic beats (S beats) and ventricular ectopic beats (V beats), the ANSI/AAMI EC57 standard suggests that classification models are also evaluated and compared by their performance on these classes. For this reason, SVEB and VEB precision, sensitivity, specificity and F1-score are reported as main performance metrics in a large portion of the research (Raj & Ray, 2018; Sellami & Hwang, 2018; Zhai et al., 2020). In addition to grouping of arrhythmias in the classes suggested by AAMI, some recent studies also perform classification on a different label set (Li et al., 2017; Qin et al., 2017; Raj & Ray, 2018). Here, the classes are more specific arrhythmia types than the ones according to the AAMI guidelines, which makes them less-represented in the datasets and the task potentially more demanding. The MIT-BIH Arrhythmia database contains up to 16 different morphological arrhythmia annotations, which are classified in 16 classes in the study in Raj and Ray (2018), achieving an accuracy of 0.963 using a Discrete Orthogonal Stockwell Transform (DOST) as a feature extraction method. Other studies, such as Li et al. (2017) and Qin et al. (2017), focus only on 6 class labels.

Various machine learning methods have been used for classification of heartbeats according to arrhythmia types, most often using standard classifiers, such as SVM and Random Forest, combined with advanced feature extraction and selection methods (Li et al., 2017; Qin et al., 2017). However, the latest best-performing methods are mostly based on deep learning. One example is the study in Mousavi and Afghah (2019), which uses a combination of a simple convolutional architecture with 3 layers, followed by an encoder–decoder architecture,

[7] https://physionet.org/.
[8] https://figshare.com/.
[9] https://zenodo.org/.
[10] https://ieee-dataport.org/.
[11] Can be obtained on USB at https://www.ecri.org/american-heart-association-ecg-database-usb.

Table 1
ECG databases for various application areas.

Database	Frequency	Leads	Length of recordings	No. of recordings (people)
MIT-BIH Normal Sinus Rhythm	360 Hz	2	Up to 24 h	18
	Applications: biometric identification			
	Studies: Elgendi (2013), Li et al. (2020), Mousavi et al. (2019), Sameni et al. (2007) and Sayadi (2008)			
MIT-BIH Arrhythmia	360 Hz	2 (MLII and V1/2/4/5)	30 min	48
	Applications: arrhythmia detection, signal denoising			
	Studies: An and Stylios (2020), Bassiouni et al. (2018), Chauhan and Vig (2015), Elgendi (2013), Hamilton (2002), Huang et al. (2019), Kiranyaz et al. (2017), Li et al. (2017), Mousavi and Afghah (2019), Qin et al. (2017), Raj and Ray (2018), Salloum and Kuo (2017), Satija et al. (2019), Sayadi (2008), Sellami and Hwang (2018), Tracey and Miller (2012), Weng et al. (2006) and Zhai et al. (2020)			
MIT-BIH Long Term	360 Hz	2 or 3	14 h–20 h	7
	Applications: arrhythmia detection			
	Studies:			
MIT-BIH Supraventricular Arrhythmia	360 Hz	2	30 min	78
	Applications: arrhythmia detection			
	Studies: Elgendi (2013)			
Sudden Cardiac Death Holter	250 Hz	2	Up to 24 h	23 (18 sinus, 1 paced and 4 AF)
	Applications: rhythmic arrhythmia detection			
	Studies: Greenwald (1986)			
MIT-BIH Atrial Fibrillation (AF)	250 Hz	2	10 h	25
	Applications: rhythmic arrhythmia detection, biometric identification			
	Studies: Li et al. (2020), Mei et al. (2018) and Mousavi et al. (2019)			
MIT-BIH ST Change	360 Hz	2	30 min	28
	Applications: biometric identification			
	Studies: Elgendi (2013) and Li et al. (2020)			
MIH-BIH Noise Stress Test (NST)	360 Hz	2	30 min	12 ECG + 3 noise
	Applications: signal denoising			
	Studies: Elgendi (2013), Huang et al. (2019), Sameni et al. (2007) and Sayadi (2008)			
Long-Term ST	250 Hz	2 or 3	21 h–24 h	86 (80)
	Applications:			
	Studies:			
Intracardiac Atrial Fibrillation	1 kHz	5 endocardial + 3 surface	1 min	8
	Applications: rhythmic arrhythmia detection			
	Studies:			
Long Term Atrial Fibrillation (AF)	128 Hz	2	24 h	84
	Applications: rhythmic arrhythmia detection			
	Studies: Elgendi (2013) and Mousavi et al. (2019)			
Atrial Fibrillation (AF) Classification - PhysioNet/CinC Challenge 2017	300 Hz	1	9 s–60 s	8,528
	Applications: rhythmic arrhythmia detection			
	Studies: Mei et al. (2018)			
European ST-T	250 Hz	2	2 h	90 (79)
	Applications:			
	Studies:			
CU Ventricular Tachyarrhythmia	250 Hz	1	8 min	35
	Applications: rhythmic arrhythmia detection			
	Studies: Nolle et al. (1986)			
BIDMC Congestive Heart Failure	250 Hz	2	20 h	15
	Applications: rhythmic arrhythmia detection			
	Studies: Baim et al. (1986)			
PTB Diagnostic ECG	1 kHz	15	10 s–2 min	549 (290)
	Applications: biometric identification			
	Studies: Fotiadou et al. (2020) and Labati et al. (2018)			
PTB-XL	1 kHz	12	10 s	21,837 (18,885)
	Applications: arrhythmia detection			
	Studies: Strodthoff et al. (2021)			
QT	250 Hz	2	15 min	100
	Applications: signal denoising			
	Studies: Elgendi (2013) and Fotiadou et al. (2020)			
St. Petersburg Institute of Cardiological Technics 12-lead Arrhythmia (INCART)	257 Hz	12	30 min	75
	Applications: arrhythmia detection			
	Studies: Elgendi (2013) and Fotiadou et al. (2020)			

(*continued on next page*)

Table 1 (*continued*).

Database	Frequency	Leads	Length of recordings	No. of recordings (people)
American Heart Association (AHA)	250 Hz	2	3 h (30 min annotated)	155
	Applications: (ventricular) arrhythmia detection Studies: Hamilton2002			
The China Physiological Signal Challenge 2018	500 Hz Applications: arrhythmia detection Studies: Yao et al. (2020)	12	6 s–60 s	6,877
Abdominal and Direct Fetal ECG (ADFE)	1 kHz	4 abdominal maternal + 1 direct fetal	10 min	5
	Applications: fetal ECG extraction Studies: Da Poian et al. (2015) and Fotiadou et al. (2020)			
Fetal electrocardiograms - B1_Pregnancy_dataset	500 Hz	4 abdominal maternal	20 min	10
	Applications: fetal ECG extraction Studies: Da Poian et al. (2015) and Fotiadou et al. (2020)			
Fetal electrocardiograms - B2_Labour_dataset	500 Hz–1 kHz	4 abdominal maternal + 1 direct fetal	5 min	12
	Applications: fetal ECG extraction Studies: Da Poian et al. (2015) and Fotiadou et al. (2020)			
Noninvasive Fetal ECG - PhysioNet/CinC Challenge 2013	1 kHz Applications: fetal ECG extraction Studies: Andreotti et al. (2014) and Da Poian et al. (2015)	1 abdominal	1 min	75 train + 100 test
Non-Invasive Fetal ECG Arrhythmia	500 Hz–1 kHz	4–5 abdominal maternal + 1 chest maternal	7–32 min	26 (14 healthy, 12 arrhythmia)
	Applications: arrhythmia detection Studies: Behar et al. (2018)			
MIMIC	1 kHz Applications: respiration extraction Studies: Charlton, Villarroel et al. (2016) and Pimentel et al. (2015)	12	/	67,830 (30,000)
Fantasia	250 Hz Applications: respiration extraction, biometric identification Studies: Elgendi (2013), Li et al. (2020) and Varon et al. (2020)	1	2 h	40
Motion Artifact Contaminated ECG	500 Hz Applications: signal denoising, exercise monitoring Studies: Behravan et al. (2015)	4	20 s	1
ECG-ID	500 Hz Applications: biometric identification Studies: Bassiouni et al. (2018) and Salloum and Kuo (2017)	1 (lead I)	20 s	310 (90)
Stress Recognition in Automobile Drivers	496 Hz Applications: emotional monitoring, respiration estimation Studies: Varon et al. (2020)	1 (MLII)	50–90 min	27
Combined measurement of ECG, Breathing and Seismocardiograms (CEB-SDB)	5kHz Applications: biometric identification, respiration estimation Studies: Li et al. (2020)	2 (leads I and II)	1 h	20
Preterm Infants Cardiorespiratory Signals (PICS)	500 Hz/250 Hz Applications: respiration estimation Studies: Gee et al. (2017)	1	20 h–70 h	10
Vortal	500 Hz Applications: respiration estimation Studies: Charlton, Bonnici et al. (2016)	1 (lead II)	10 min	45
ECG-Fitness	/ Applications: signal denoising, exercise monitoring Studies: Spetlik et al. (2018)	2	1–2 min per activity	17
Glasgow University Database	250 Hz	3 (leads II & III and cheststrap V2-V1)	10 min (2 min per 5 activities)	25
	Applications: signal denoising, exercise monitoring Studies: Porr and Howell (2019)			
Shaoxing People's Hospital's 10,000 patients arrhythmia database	500 Hz Applications: arrhythmia detection Studies: Yildirim et al. (2020)	12	10 s	10,646

(*continued on next page*)

Table 1 (*continued*).

Database	Frequency	Leads	Length of recordings	No. of recordings (people)
Continuously Annotated Signals of Emotion (CASE)	1 kHz	2 or 3	40 min (a few min. per video)	30
	Applications: emotional monitoring Studies: Zhang et al. (2021)			
Cognitive Load, Affect and Stress (CLAS)	256 Hz Applications: emotional monitoring Studies:	1	30 min	62
Intelligent Athlete Monitoring for Cardiovascular Wellness (iAMwell)	2 kHz	2 or 3	/	15 (6 athlete and 9 non athlete)
	Applications: signal denoising, exercise monitoring Studies:			
MARSH	/ Applications: respiration estimation Studies: Pirhonen and Vehkaoja (2020a)	2 (lead I and II)	15 min	29
Atrial Fibrillation (AF) Screening	300 Hz Applications: rhythmic arrhythmia detection Studies:	1 (lead I)	40 s	2422
Intercity Digital Electrocardiogram Alliance (IDEAL)	200 Hz Applications: arrhythmia detection, biometric identification Studies: Labati et al. (2018)	3	24 h	202
Lobachevsky University Electrocardiography (LUDB)	500 Hz Applications: arrhythmia detection Studies:	12	10 s	200

built from bidirectional LSTM units. They also perform oversampling of the less-represented classes using the SMOTE technique, as a solution to one of the most challenging aspects of the MIT-BIH Arrhythmia database — class imbalance. This problem is also successfully addressed in Sellami and Hwang (2018), where a novel loss function for a convolutional neural network is proposed. This function, which changes the loss weights dynamically according to the distribution of classes in each batch, is referred to as batch-weighted loss. Using batch-weighted loss, an aggregate F1 score of 0.897 is achieved for the inter-patient evaluation schema for all 5 classes, which is the state-of-the art result for this task on the MIT-BIH Arrhythmia database.

All of the above mentioned methods require heartbeat segmentation because the learning algorithms operate on a single-heartbeat level. A slightly modified approach is employed when using the new generation of datasets mentioned previously, which contain sequence-level annotations. Heartbeat form classification has been successfully performed using different types of neural network architectures, most notably using an improved residual neural network (xresnet) and achieving an AUC of 0.896 (Strodthoff et al., 2021). They use an overlapping sliding window approach by training the classifier on 2.5 second-long segments and generating a prediction for each of the segments. These predictions are then aggregated to produce a single prediction for the entire signal. This study demonstrated that segmenting the heartbeats is not necessary to obtain satisfactory arrhythmia classification performance.

Another direction of research in arrhythmia detection is utilizing all 12 leads present in standard clinical ECG. There have been many new interesting datasets containing 12-lead ECG recordings suitable for this purpose, such as the 2018 China Physiological Signal Challenge Dataset (Liu et al., 2018), the smaller but highly annotated Lobachevsky University Electrocardiography Database (LUDB) (Kalyakulina et al., 2020) with 200 recordings, and the previously mentioned PTB-XL database and the Shaoxing People's Hospital's 10,000 patients arrhythmia database, which are of great significance due to their size and contain heartbeat form labels, in addition to rhythm labels and other diagnostic statements. These recordings are generally shorter that those obtained with a Holter monitor, ranging from 10 s up to 1 min in most databases. This kind of data allows for fusion of all 12 ECG leads in deep learning architectures in order to obtain more reliable arrhythmia classification. Classification accuracy of 81.2% has been achieved on 8 classes (consisting of both morphology and rhythm abnormalities) present in the 2018 China Physiological Signal Challenge Dataset in Yao et al. (2020). This study proposes attention-based time-incremental CNN that successfully fuses the information from all leads, both spatially, using convolutional layers, and temporally, using LSTM units. Furthermore, the attention module on top of all other layers enables for the network to concentrate on the informative parts of the signal and additionally makes the model more interpretable. The winning submission of the challenge (Chen et al., 2020) proposes an interesting classification approach to utilize all 12 leads, which is also based on CNNs, but is enhanced with an ensemble model combining 12- and 1-lead models.

The research in the area of detection of morphological arrhythmias has been almost exclusively focused on classifying the heartbeats in the MIT-BIH Arrhythmia database in the 5 groups of arrhythmias established by AAMI. Some studies report almost perfect results for this specific problem, for example, the study in Acharya et al. (2017) reports overall precision and recall of around 96%–97%. However, this is achieved using the intra-patient evaluation paradigm and does not reflect a realistic scenario. Due to this variability in the evaluation procedures employed, some of which are highly flawed, as shown in Section 2.5, as well as the limited number of test subjects in this public database, there is still a need for further research before employing automatic machine learning models for detecting arrhythmias in clinical practice. Standardization of the evaluation procedure, as well as including representative heartbeats from a variety of data sources, instead of only one database, is necessary to further advance the research area of heartbeat classification for arrhythmia detection.

3.2. Rhythm detection and classification

Rhythmic arrhythmias are characterized by a set of irregular heartbeats. The most common type is atrial fibrillation (AF) and the most widely used databases that contain such recordings are MIT-BIH AF (Moody & Mark, 1983), the dataset from the PhysioNet Computing in Cardiology (CinC) Challenge 2017 (Clifford et al., 2017), and Long Term AF (Petrutiu et al., 2007). In addition, the new generation of public ECG datasets in the last few years, including the 2018 China Physiological Signal Challenge Dataset (Liu et al., 2018), the PTB-XL database (Wagner et al., 2020) and the Shaoxing People's Hospital's 10,000 patients arrhythmia database (Zheng et al., 2020), covers a wide range of rhythms in addition to atrial fibrillation, including different types of tachycardia, bradycardia and heart blocks.

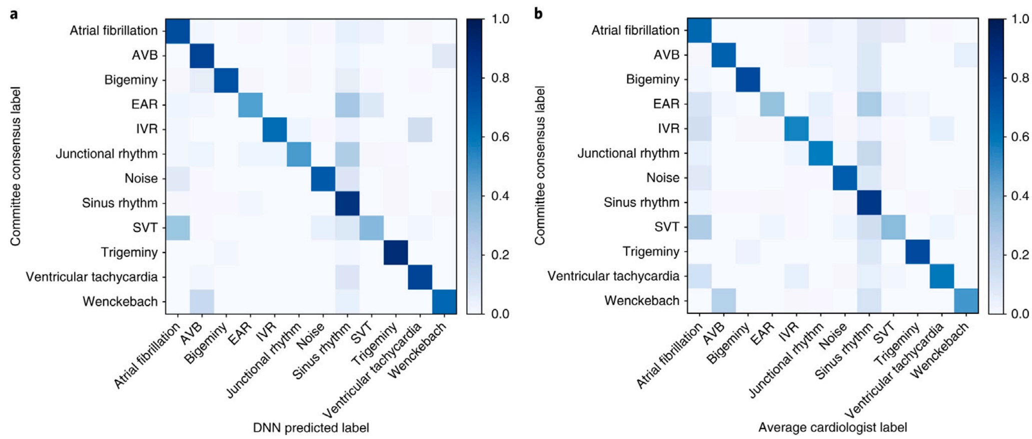

Fig. 5. Confusion matrix from 12-class arrhythmia classification performed by (a) DNN model and (b) cardiologist in the Stanford study (Hannun et al., 2019).

The MIT-BIH AF database contains short recordings, mostly 30 s long, and they belong to one of 4 categories: atrial fibrillation, normal rhythm, noise, and other rhythm. The Long Term AF database, on the other hand, offers 24-h long recordings and a larger number of annotations (such as atrial bigeminy, supraventricular tachyarrhythmia, sinus bradycardia, etc.). Classification of such rhythmic arrhythmias has been successfully done by analyzing heart rate variability, as feature in standard machine learning algorithms (Mei et al., 2018), as well as by employing mechanisms in deep neural networks that capture the temporal characteristic of ECG signals, such as attention and LSTM neural networks. A state-of-the-art attention-based deep neural network (Mousavi et al., 2019) has achieved one of the highest performances at AF detection, with F1-score of 0.994 on the MIT-BIH AF database, for two-class classification (AF and Non-AF rhythm). This network architecture consists of two channels: the first one is the raw ECG data input into attention and LSTM layers, while the second one is a 2-dimensional representation of the wavelet power spectrum of the signal as an input in LSTM layers. The Shaoxing People's Hospital's 10,000 patients arrhythmia database has been used to develop a rhythm classification model (Yildirim et al., 2020) using a convolutional neural network and achieving an accuracy of 0.9224 for 7 rhythm types.

There are a few recent successful works on heartbeat classification utilizing data that is not publicly available (yet). These studies mostly focus on rhythmic arrhythmia and include a wider range of different arrhythmia labels. One of them is the dataset used in the study by the Stanford Machine Learning (ML) Group (Hannun et al., 2019). They collected the data using a single-lead mobile ECG sensor, the Zio Patch (Turakhia et al., 2013). Almost 30,000 patients were continuously recorded for up to 14 days. Cardiologist-level performance (F1-score of 0.837) in arrhythmia detection for 12 different rhythm abnormalities was achieved using a deep learning approach. The network architecture used is a very deep residual network — 34 convolutional layers are grouped in blocks with skip connections. The final classification results of this study are given in Fig. 5, summarized in a confusion matrix. It can be observed that the deep neural network model performs comparable to a trained cardiologist. Another such novel dataset is included in the Clinical Outcomes in Digital Electrocardiology (CODE) study (Ribeiro et al., 2020), which consists of 2 million short 12-lead clinical ECG recordings, including AF, sinus bradycardia and tachycardia. They exhibit promising results (F1 score of 0.925) using a deep learning network with residual blocks, similar to the one in the Stanford study.

Atrial fibrillation is the most wide-spread type of arrhythmia. Therefore, it is understandable that the most important recent advances in arrhythmia detection from ECG are mostly focusing on atrial fibrillation and distinguishing it from other common heart rhythms. Probably the most noteworthy ECG analysis study in general is the study performed by the Stanford ML Group, because of the size and the scope of the data they collected, but also because the recordings come from a mobile wireless ECG device, which is still very rare in this research field.

3.3. Signal quality assessment

The quality of the ECG signal is crucial for successful ECG analysis, which is why a lot of studies focus on its assessment and improvement. When developing a novel ECG measurement scenario or device, it is important to be aware of the quality of the signal obtained in order to have information on what kind of analysis could be done with such measurements. Also, a single low-quality recording included in a dataset could influence the performance of the entire method.

Signal quality assessment and improvement is often done as one of the first steps when performing ECG analysis for any purpose, as described in Section 2.1. In addition, in recent years, it has been shown that this can also be performed using machine learning and can be formulated as a classification task ("acceptable" vs. "unacceptable") (Satija et al., 2019). The standard approach, however, is to focus on improving the quality of the signal with denoising techniques (An & Stylios, 2020; Huang et al., 2019; Sameni et al., 2007; Sayadi, 2008; Tracey & Miller, 2012; Weng et al., 2006). Most of these techniques are based on signal processing, the most notable of which is the modification of the Savitzky–Golay (SG) filter in Huang et al. (2019). The SG-filter variation, called low-distortion adaptive Savitzky–Golay filter (LDASG), is adaptive to the signal variations and is based on discrete curvature estimation of the signal, whereby reducing the distortion introduced by the filter. Interesting visual examples taken from this study are shown in Fig. 6, where the denoising abilities of different techniques can be observed when applied to noisy recordings with different signal-to-noise ratios (SNRs), namely SNR = 0 db and SNR = 10 dB. When compared to the presented NLM and EMD-wavelet methods, the proposed LDASG filter shows very good results, successfully reducing the noise, whilst keeping all ECG characteristics. Namely, the average mean squared error (MSE) for the LDASG filter is lower by 33.33% when compared to EMD-wavelet and by 50% when compared to NLM, while the average Percent Root mean-square Difference (PRD) is lower by 18.25% when compared to EMD-wavelet and by 25.24% when compared to NLM.

Deep learning has also shown promising results in recent efforts, usually by using some type of encoder–decoder architecture, such as the fully convolutional encoder–decoder with skip connections in Fotiadou et al. (2020), inspired by a model developed for image restoration.

The standard evaluation procedure for denoising techniques is to simulate noisy ECG by adding noise to clean ECG. This procedure was used to obtain the MIT-BIH Noise Stress Test (NST) database (Moody

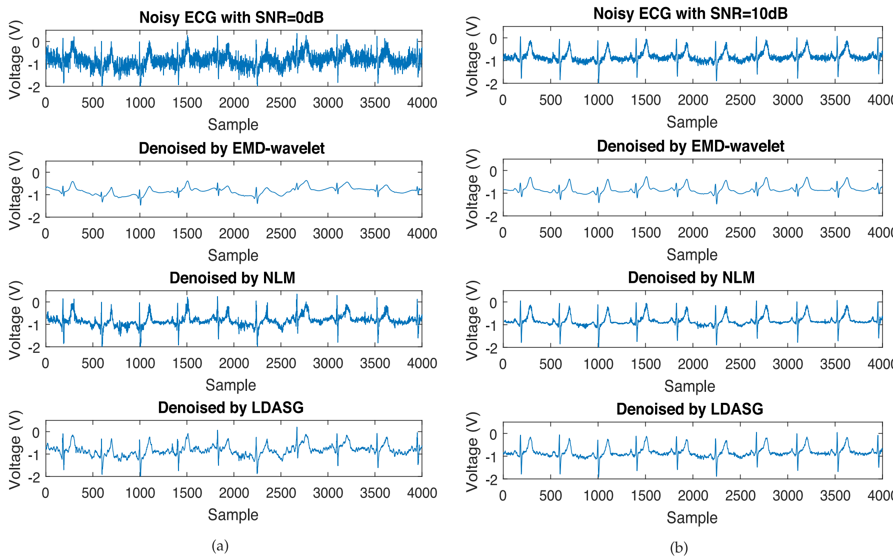

Fig. 6. ECG denoising using EMD-wavelet, NLM and LDASG filter (Huang et al., 2019). (a) Results on noisy ECG with SNR of 0 db; (b) Results on noisy ECG with SNR of 10 db.

et al., 1984) – by adding noise recorded during a stress test to clean recordings from the MIT-BIH Arrhythmia database (Moody & Mark, 2001). According to the AAMI EC57 and the IEC 60601-2-47 standards, performance evaluation on the NST database is a must when testing an ECG denoising technique. The datasets used in Fotiadou et al. (2020), Huang et al. (2019) and Satija et al. (2019) are also generated in a similar way. In these studies, either White Gaussian Noise (WGN), nonstationary real Muscle Artifacts (MA) noise or real recorded electrode noise is added to clean recordings from MIT-BIH Arrhythmia, PTB (Bousseljot et al., 1995), INCART (Tihonenko et al., 2007) or QT (Laguna et al., 1997) databases. In this way, semi-realistic noisy ECG recordings are created, which usually correspond to the real types of noisy ECG encountered during recording situations. For quantitative evaluation of a denoising technique, the average performance metrics (MSE, PRD, SNR improvement) are calculated on noisy ECG signals at different SNR levels.

A lot of recent studies prefer to use realistic noisy ECGs and start to focus on recording ECG using a wireless sensor during different physical activities. In that case, quality assessment needs to be applied to those novel datasets, usually small in size, such as the data in Ilic et al. (2019) collected with the Savvy sensor. There is a trend in recent years to make these smaller ECG datasets public. Examples thereof are the iAmWell Dataset (Elia et al., 2017), including recordings during different types of running, and the Wearable Ambulatory ECG Dataset (Kher, 2020), containing records with only simple activities, such as raising the arms and sitting down/standing up. However, the most notable one is the Glasgow University Database (Porr & Howell, 2019), where it has been shown that the realistic movement noise introduced in ECG recordings significantly influences the performance of most QRS detection algorithms. They also highlight additional shortcomings of standard databases, such as the MIT-BIH Arrhythmia, most notably imprecise R-peak annotation locations, and how this results in non-realistic R-peak detection accuracy reported in research so far. This further confirms the need for new, better and more realistic public ECG data, most importantly data that contains realistic movement noise. Furthermore, there have been some very recent efforts which attempt to assess the quality of the signal by quantifying how well a specific QRS detector can interpret the ECG signal (Chromik et al., 2021). Here, the quality assessment is not independent of the other ECG tasks, but is connected to the QRS detection step and is measured by a proposed certainty metric for QRS detection, which could also serve as a signal quality metric.

ECG measurement standards, most commonly the 12-lead ECG and the Holter monitor, have proven to acquire recordings of sufficient quality for the medical analyses, usually performed in clinical practice. Novel measurement technologies, however, as well as the modern measurement scenarios they aim to be used in, require new and more extensive research examining the quality of the measurements acquired, as well as how standard ECG algorithms (denoising, QRS detection) can be applied to them. For this reason, the publication of data recorded using novel ECG measuring devices is of extreme importance.

3.4. Biometric identification

Biometric recognition is a mature field of research. However, the use of physiological signal features for this task, such as the ECG trails, still needs further improvements, although it has shown promising results. There are still some challenges to overcome in order to make the ECG a widely accepted biometric identity, some of which are the questions of uniqueness and permanence over time of the ECG markers (Carreiras et al., 2014). Concerning uniqueness, it needs to be confirmed that ECG from one person is significantly different from the ECG of all other people. Permanence over time, on the other hand, requires that a person's ECG does not change over some period of time, depending on the use-case.

A lot of studies include data specifically collected for the task of biometric identification. There are studies that use less-invasive ECG recording methods than those used in clinical settings, like mobile ECG devices (Agrafioti & Hatzinakos, 2010; Šprager et al., 2017), however, the recordings are still not publicly available. Other recent efforts include using ECG databases made specifically for the purpose of identification, such as the ECG-ID database (Lugovaya, 2011) in Bassiouni et al. (2018) and Salloum and Kuo (2017). More recently, other general ECG databases, collected for other tasks, have been used also for the task of biometric identification, such as: the healthy subjects subset (with identifier E-HOL-03-0202-003) of the IDEAL database (Couderc et al., 2005), PTB (Bousseljot et al., 1995) in Labati et al. (2018), Fanstasia (Iyengar et al., 1996), CEB-SDB, Normal Sinus Rhythm database (Moody, 1999), MIT-BIH ST Change database (Albrecht, 1992) and MIT-BIH AF (Clifford et al., 2017) in Li et al. (2020). However, most often used are the ECG-ID and the PTB datasets. ECG-ID offers short 20-second long recordings with 10 annotated beats (QRS complexes) from 90 healthy volunteers. There are at least 2 and up to 20 different recordings for one person. The PTB database, on another hand, is a general diagnostic ECG database and includes subjects with various diagnoses and their medical history, in addition to short (up to 2 min) 16-lead ECG recordings.

Two main tasks are usually addressed in the biometrics studies: authentication and identification. The more challenging of the two is identification, which involves determining the identity of a person, using its biometric features/markers, given information about the features/markers of all possible identities. For this task, statistical feature extraction methods combined with simple classifiers have shown very successful results so far, for instance, higher-order statistics (HOS) (Šprager et al., 2017) and linear discriminant analysis (LDA) (Agrafioti & Hatzinakos, 2010). A deep neural network can also be employed as a feature extractor, for example, CNN (Labati et al., 2018). With this approach, a 100% testing accuracy has been achieved on the PTB database for closed-set identification. However, both ECG-ID and PTB databases lack some aspects which do not allow these results to be applicable in wide daily practice. ECG-ID slightly covers the problem of confirming permanence over time of ECG markers, by employing multiple measurements of the same person from up to 6 month periods, however, the number of participants is fairly small (90), which is not enough to confirm uniqueness. PTB, on another hand, offers a larger number of people (290), but there is only one short recording for many of the subjects, leaving open questions about the permanence over time of the biometric ECG features.

ECG for biometrics offers advantages over other methods, mostly in terms of its inimitability, but it still has not been applied in practice, even though various studies have reported very high identification scores on different datasets. The reason for this is the small scope of these studies, however, wider research should elevate the promotion of the application of ECG methods for biometric identification in practice in near future.

3.5. Respiration estimation

ECG signals are a result of the electrical activity of the heart. However, there are other processes which also influence this signal, such as muscular activity and respiration. Respiration is one of the most informative vital signs for the physiological state of a person or the progression of an illness. Its continuous monitoring is valuable not only in hospital environments, but also during daily activities. By utilizing the phenomena that respiration modulates the ECG signal, respiration can be extracted from it without using any additional sensors, which supports the idea of a multi-functional device for monitoring various vital signals by combining a minimal number of body sensors with different functions into a single one (Depolli et al., 2016; Trobec et al., 2014).

The challenges in this application area include estimating the respiratory signal and locating the breathing cycles. There are various methods for these tasks, but they are not deterministic. Consequently, there is a variety into the methods and results, both in the step of estimating the respiratory waveform, as well as in locating the respiratory cycles, which makes various studies difficult to compare. It should be also noted that various performance metrics are used in research to evaluate the methods, e.g. limits of agreement (95% LOA), bias and 2 standard deviations (2SD) in Charlton, Bonnici et al. (2016), while Varon et al. (2020) used relative error for evaluating respiratory rate estimation, and cross-correlation and spectral coherency for evaluating wave morphology similarity.

From the literature search, it can be concluded that most methods perform signal processing to transform the ECG signal into a respiratory one (Charlton, Bonnici et al., 2016; Sohn et al., 2017; Trobec et al., 2012; Varon et al., 2020). Some of these studies use their private data specifically collected for this purpose, however, there are a few public datasets suitable for testing methods to derive respiration from ECG, such as MIMIC, Fantasia and Drivers dataset (Healey & Picard, 2005) in Varon et al. (2020). In Fantasia and MIMIC, the subjects are recorded during resting, in the intensive care in the case of MIMIC and while watching a movie in Fantasia, while the Drivers dataset includes different levels of stress during driving, which results in various respiratory rates present in the recordings. The data from the very recently published MARSH study (Pirhonen & Vehkaoja, 2020a, 2020b) also belongs to this group and it consists of simultaneous ECG, PPG and respiration signals, including various controlled respiratory rates of the subjects ranging from 0.1 Hz up to 0.6 Hz. In addition, the respiration cycles are annotated, which makes this data suitable for respiration extraction experiments.

The most comprehensive study of signal processing methods for estimation of respiration from ECG is Charlton, Bonnici et al. (2016), which uses the VORTAL dataset, collected for the purpose of this study and openly available on request. The dataset comprises of simultaneous recordings of both young and elderly subjects during lying supine as well as during physical activities, such as walking and running from the young subjects. This study implemented 314 combinations of techniques for estimation of both respiratory signal and respiratory rate, including both feature- and filter-based algorithms. The conclusion from the study is that the best performing method for estimation of respiratory signal is feature-based operating on ECG by fusing techniques such as baseline wander (BW), amplitude modulation (AM) and frequency modulation (FM). When it comes to estimating the respiratory rate, time-domain breath detection was shown as a better solution than frequency-domain. The measurements from the best performing method, a fusion of feature-based ECG-derived respiration (EDR) and time-domain breath detection, had a bias of 0.0 bpm (breaths per minute) and 95% LOA of – 4.7 bpm to 4.7 bpm. Impedance pneumography (IP) – the clinical standard for continuous monitoring of respiratory rate in a hospital environment – was ranked 5th according to performance (after 4 ECG-derived algorithms), with a bias of −0.2 bpm, 95% LOA of – 5.6 bpm to 5.2 bpm. This suggests that ECG-based algorithms could be sufficiently precise for use in clinical practice (Charlton, Bonnici et al., 2016). Other studies also make use of the amplitude modulations of the ECG signal by respiration and sometimes use data collected for the specific purpose of respiration estimation, which is not publicly available yet (Trobec et al., 2012).

Besides ECG, the photoplethysmogram (PPG) waveform has also been used to estimate respiration. PPG is another way to track the activity of the heart — by measuring changes in blood volume over time. However, the signal obtained in this way is less informative of the state of the heart and is usually used only to estimate the heart rate. The measurement procedure, on another hand, is simpler than obtaining ECG. It involves a wrist-worn devices, which is why a lot of studies focus on respiration extraction from PPG. For this purpose, many of the mentioned datasets, such as MIMIC and Vortal, besides ECG and respiration, contain also simultaneous PPG recordings. It has been shown that PPG can also be used for respiration estimation (Charlton et al., 2018), but the algorithms examined in Charlton, Bonnici et al. (2016) generally performed better when applied to ECG, with a median decrease in the 2SD of 0.8 bpm when compared to PPG. Respiratory waveforms obtained from ECG are more precise than those from PPG, which shows that PPG is not a replacement for ECG.

The area of EDR estimation is not so well-defined when compared to, for example, arrhythmia detection. When it comes to evaluation methods and data used, there are not many standard practices employed by various studies that allow for their comparison. However, the studies have shown that, in general, a reliable respiration waveform can be extracted from an ECG signal. It should also be noted that respiration estimation techniques are usually based on signal processing and there have not been yet studies using deep learning techniques for this purpose, as opposed to other application areas where these methods are well established. For this reason, a possible direction for future work in ECG-derived respiration could be the use of neural networks.

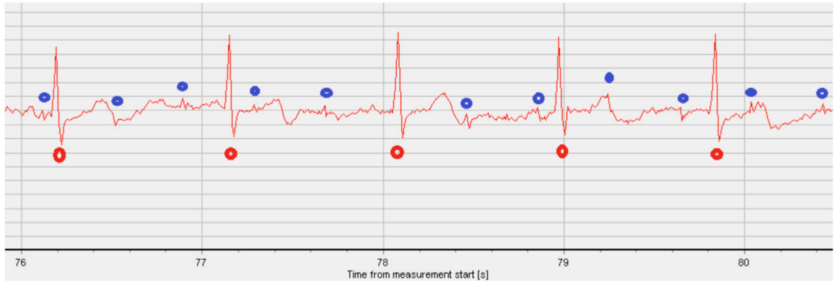

Fig. 7. Annotated maternal (red markers) and fetal (blue markers) beats in abdominal ECG (Rashkovska & Avbelj, 2017).

3.6. Fetal ECG extraction

An interesting demonstration of the multi-functionality of ECG body sensors is their use for monitoring fetal ECG (FECG). This can be achieved through recording abdominal ECG (AECG) by placing the electrodes on the maternal abdomen. It is a non-invasive method for monitoring the cardiac activity of a fetus and has the potential to be preferred over other more invasive methods. An example of such a signal, containing both maternal and FECG components, with annotated heartbeat locations, can be observed in Fig. 7, taken from the study in Rashkovska and Avbelj (2017). The challenge in this tasks is to successfully extract the FECG signal from the AECG.

There are public databases of abdominal ECG, such as Abdominal and Direct Fetal ECG (ADFE) database (Jezewski et al., 2012) and Non-Invasive Fetal ECG Arrhythmia database (Behar et al., 2018), available on PhysioNet. These databases are joined in the 2013 PhysioNet/CinC Challenge dataset (Goldberger et al., 2000; Silva et al., 2013), together with additional both private and simulated FECG sources. This challenge dataset is most widely used in fetal ECG extraction studies, in addition to the ADFE database separately. The ADFE database contains signals gathered during labor from 5 different women. It consists of four differential signals obtained from the maternal abdomen and a reference direct FECG, recorded simultaneously using a spiral electrode attached to the fetal head. The entire 2013 PhysioNet/CinC challenge data contains 75 AECG recordings with annotated fetal heartbeats, covering only normal rhythm, as well as 100 test recordings without annotations. It has been concluded that the separation of FECG is best done using ICA, shown both on the ADFE database (Da Poian et al., 2015; Fotiadou et al., 2020) and the 2013 PhysioNet/CinC challenge data (Andreotti et al., 2014). This method successfully separates the maternal and FECG waveforms as independent components of the signal. Another approach (Rashkovska & Avbelj, 2017) applies expert knowledge about the size of the heart of the fetus and its QRS amplitude compared to the mother's, to separate fetal heartbeats, using data specifically collected for this study with the Savvy sensor.

Similar to respiration estimation, there are various ways to evaluate the methods for FECG estimation since there is no unambiguous correct FECG as opposed to, for example, classification tasks such as arrhythmia detection. A simpler solution is to only perceive fetal heart rate (fHR) and calculate mean squared error (MSE) between the estimated and true fHR as a performance metric (Clifford et al., 2011; Da Poian et al., 2015; Silva et al., 2013). However, methods attempting to evaluate the entire FECG signal, including heartbeat locations, range from comparing characteristic ECG intervals between estimated and true FECG waveforms, e.g. QT and PR intervals (Fotiadou et al., 2020), RR interval (Da Poian et al., 2015), to calculating classification metrics (sensitivity, specificity, precision, accuracy and F1-score) of estimated QRS complex locations, as defined in the ANSI/AAMI EC57 standard (ANSI/AAMI EC57, 2012). The standard allows for ±50 ms acceptance interval between the detected and the closest reference annotation. In this way, each fetal heartbeat can be effectively regarded as either correct or false. This type of evaluation has been performed on both the ADFE dataset (Da Poian et al., 2015) and the 2013 PhysioNet/CinC challenge data (Andreotti et al., 2014; Da Poian et al., 2015). A sensitivity of 92.5% for the ADFE dataset is obtained by applying ICA to compressed abdominal FECG (Da Poian et al., 2015). On the 2013 PhysioNet/CinC challenge data, average sensitivity of 97.4% was achieved (Andreotti et al., 2014). An example reconstructed FECG waveform from the 2013 Challenge dataset can be observed in Fig. 8, taken from the study in Da Poian et al. (2015).

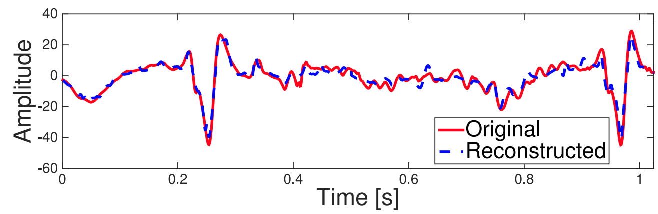

Fig. 8. Ground truth and reconstructed fetal ECG waveforms (Da Poian et al., 2015).

An extended dataset containing abdominal and direct fetal ECG has been published very recently (Matonia et al., 2020), where the well-established ADFE database available on PhysioNet is a small portion of this collection. This new dataset consists of two subsets: one subset (B2_Labour_dataset) is an extension of the ADFE database and contains 12 recordings during labor with reference direct FECG, while the second subset (B1_Pregnancy_dataset) is recorded earlier during pregnancy from 10 women with expert annotations of fetal heartbeat locations. This database offers more comprehensive recordings than all previous public non-invasive abdominal fetal ECG datasets, which makes it a valuable addition and is expected to be used in future studies.

Using expert-annotated fetal ECG data, it has been shown that fetal heart activity can be accurately estimated from AECG. Furthermore, application examples prove that the estimation of FECG from AECG is feasible, even when ground truth signal (simultaneous FECG) is unavailable. Such studies (Rashkovska & Avbelj, 2017), performed on recordings other than the ADFE, have shown consistency with medical knowledge about FECG, which confirms that abdominal ECG is an accurate non-invasive method for monitoring fetal cardiac activity.

3.7. Physical and emotional monitoring

In addition to the areas covered so far, ECG has the potential to be used for monitoring the physical state and activity of a person as well as his/her emotions. Such tasks can only be accomplished when combining ECG with other physiological signals from other types of body sensors (e.g. inertial sensors for physical activity, galvanic skin response for emotions).

One important application scenario is in monitoring athletes during demanding physical exercise and it is of high importance that the signal obtained is of sufficient quality (Ilic et al., 2019) – area more extensively elaborated in Section 3.3. It has been also determined that ECG measuring devices on the chest are of great importance and cannot be replaced with different wrist-worn alternatives and PPG (Seshadri et al., 2019). There are a few smaller publicly available datasets that contain measurements during different physical activities, such as the Glasgow University Database (Howell & Porr, 2018), Wearable Ambulatory ECG (Kher, 2020) and iAmWell (Elia et al., 2017). The scope of these databases is limited, generally including only a few activities and shorter recordings. The iAmWell dataset includes data from both athletes and non-athletes, while the Glasgow University Database includes, besides ECG, also simultaneous accelerometer recordings, both of which represent recording scenarios suitable for subsequent exercise monitoring.

The situation is similar when it comes to monitoring different human emotions using ECG. Few datasets exist such as CLAS (Markova et al., 2019), CASE (Sharma et al., 2019) and Drivers Dataset (Healey & Picard, 2005), which contain ECG measurements during various emotional states of a person. The CLAS and CASE datasets record different emotional responses (e.g. boring, scary) from subjects by playing videos with diverse content as well as by setting tasks for the subjects (e.g. math problems). However, the most significant task in the field of emotional monitoring is stress detection (Iqbal et al., 2021), especially in the modern world. For this purpose, the Drivers Dataset has been widely used. The entire Drivers Dataset is recorded during driving, but through different parts of the city, which was presumed to result in various specific levels of stress. What is common for these datasets is that multiple signals are recorded at the same time, e.g. respiration, galvanic skin response, acceleration, because only the combination of sensors can allow for successful monitoring. Excluding the Drivers Dataset, both CLAS and CASE datasets have been made public very recently, confirming the importance of this direction of research in the future, with ECG signals playing a significant role.

The potential range of ECG monitoring is wider than the already well-established use for diagnosing heart conditions. Emotional and physical monitoring using ECG are two noteworthy directions in research proving this potential. However, the data and analyses done in this direction so far have not been very extensive, usually focusing on one part, such as QRS detection, while end-to-end analysis has not been performed yet.

3.8. Commercial ECG analysis software

ECG analysis is a field which has been researched for over 40 years. During this time, commercial software products have been developed and have become well-established for some tasks, e.g. QRS detection and ECG morphology features extraction. One of the most notable ones is the University of Glasgow (Uni-G) ECG Analysis Software, described in Macfarlane et al. (2005). The algorithms included in this software have been marketed by multiple companies such as Cardioline,[12] Spacelabs Healthcare[13] and AMPS,[14] where their product Continuous ECG Recording Suite (CER-S) is one of the most successful commercial ECG interpretation software. It includes multiple modules, including detection of heartbeat locations, generation of template beats, and beat measures (QRS duration and amplitude, ST slope and duration, etc.). Furthermore, the software can perform rule-based diagnostic interpretation and classification, including detection of isolated beats, couplets, bigeminy, and trigeminy, and rhythm analysis with detection of tachycardia, bradycardia and atrial fibrillation.

In addition, software tools, such as Uni-G, are being used in research studies in the field of medicine (Demidova et al., 2019), as well as during the initial handling of ECG data when the goal is a more complex automatic task. This is the case in the CODE study (Ribeiro et al., 2020), where the Uni-G software is used as a complementary tool in a few parts of the processing pipeline in the development of a novel deep learning method for arrhythmia detection.

There are other widely-used software solutions, such as the Hanover ECG System (HES) (Zywietz et al., 1990) evaluated in Khawaja et al. (2011), offering similar services as the Uni-G software. In addition, HES offers different software modules for resting ECG, exercise ECG and real-time analysis, handling the ECGs obtained in these modes in slightly different ways (e.g. more focus on denoising in exercise mode). Traditionally, these platforms perform only the first few steps of the computational analysis pipeline: denoising, segmentation (by detecting beat locations), and some kind of feature extraction by calculating beat measures. However, some of these software tools also incorporate rule-based diagnostic classification techniques, such as the ST-Elevation Myocardial Infarction (STEMI) detection criteria (Macfarlane et al., 2004), included in the Uni-G software (Macfarlane et al., 2005), and the ventricular beat detection pipeline, implemented in the open-source EP Limited software,[15] as described in Hamilton (2002).

In recent years, the research advances towards more complex machine learning techniques for ECG analysis, which are slowly starting to be integrated in commercial ECG analysis software as well. Examples of this are deep learning based platforms, such as Cardiomatics[16] and Cardiologs,[17] which are able to classify ECG segments into a wide range of common arrhythmia types, such as AF and ventricular beats (Smith, Rapin et al., 2019; Smith, Walsh et al., 2019). Both Cardiomatics and Cardiologs are implemented as cloud-based services, aimed for postponed analysis of ECG, and are intended for ECG measurements obtained with Holter monitors. On another hand, older traditional software solutions (Uni-G and Hanover) are usually implemented on the ECG measuring device itself.

Wireless ECG devices are the latest novelty in ECG measurement technologies. They are incorporated within ECG analysis platforms, both for postponed analysis, as well as real-time monitoring. Zio patches[18] are extensively used for daily ECG monitoring, both to obtain a long-term measurement, which is later analyzed by a cardiologist with ZioXT, as well as continuous telemetry monitoring for high-risk patients with ZioAT. Other such examples are various devices by Cardionet and Lifewatch,[19] where depending on the patient's needs, both real-time monitoring with irregular heartbeat alerts, as well as postponed analysis can be performed. Similar services are offered by Cardiolyse[20] and AliveCor with Kardiacare,[21] usually in combination with a smartphone application that displays the ECG features.

4. Discussion

In this section, discussion is made regarding the three main aspects of the computational ECG analysis, surveyed in this paper: methods, data, and applications. First, we start with the challenges set by the penetration of mobile ECG devices.

4.1. Challenges for mobile ECG devices

The Holter monitor has been a standard for long-term ECG monitoring for over 50 years. Future research in this field, however, will expand the use of ECG analysis in many different areas, all thanks to wireless mobile ECG body sensors. These new devices bring new challenges in this field. Some of them are related to the device itself, such as lowering the sampling frequency for longer power autonomy and reducing the number of ECG leads measured (usually one lead) in order to be more

12 https:www.cardioline.it/en/home-eng/.
13 https:www.spacelabshealthcare.com/.
14 http://www.amps-llc.com/prodotti-holter.
15 https://www.eplimited.com.
16 https://cardiomatics.com/.
17 https://cardiologs.com.
18 https://www.irhythmtech.com.
19 https://www.gobio.com.
20 https://cardiolyse.com/.
21 https://store.kardia.com/products/kardiacare.

unobtrusive to the user. Also, since communication protocols take up a large part of the total power consumption in most portable wireless devices (Altini et al., 2011), dedicated communication protocols need to be developed for wireless data transmission of ECG data, such as the PCARD wireless protocol in Depolli et al. (2016). Moreover, large volumes of ECG data are being recorded using these new measurement devices. Due to this, compression is another very significant topic in computational ECG analysis, most notably in the case of wireless ECG devices. Furthermore, the processing power and storage capacity on the device are limited, which additionally limits the implementation of algorithms on the device itself, and requires dedicated methods that would take these limitations into account (Marsili et al., 2020).

Other challenges are closely related to the use-cases of the sensors. The placement of such sensor could vary with each use, as it is not (always) placed by a trained professional, as in the case of Holter monitors. In addition, long-term monitoring includes recording during all kinds of movements of the subject, which adds noise and disturbance to the ECG. This results in the need for more robust methods, as well as more versatile data to develop these methods. In the following two sections, a critical review of the state-of-the-art will be provided, in relation to how they answer to these challenges and needs.

4.2. ECG analysis methods

The paper so far gave an overview of the current trends in the field of computational ECG analysis. The focus is on the applied methods and used data in various studies from different areas of application. In the first part, we focus on the steps in the ECG analysis pipeline and we can observe that the preprocessing methods used are fairly standard and consistent across different studies. These methods include signal processing techniques that are already well-established. However, with the more frequent use of long-term daily activity ECG measurements and making use of the multi-functionality of body sensors, signal quality and denoising have an increasingly important role and advances are being made (Huang et al., 2019). Very recent studies, such as Fotiadou et al. (2020), have attempted to perform quality improvement on a complex signal, such as the extracted fetal ECG, using a deep encoder–decoder architecture. The main challenge still remains the denoising of muscle noise and artifacts that have frequency spectra overlapping with the ECG spectra, especially in measurements acquired with mobile ECG devices. Moreover, there is a lack of studies that deal with signal quality assessment as a classification task. The same goes for QRS detection algorithms: while there are well-established methods for their current use, mobile ECG devices, and the new challenges they bring, will require more accurate and faster QRS detection, which is why studies developing such methods, like (Elgendi, 2013), are of great importance.

When looking at the highest-performing newer methods in all application areas, another remark concerning the segmentation step can be made in general. They mostly use some type of a sliding window technique, instead of the beat-by-beat classification dominant in more traditional older studies. This variant is also more simple and less prone to preprocessing errors due to the elimination of the need for QRS detection. Regarding feature extraction methods, a conclusion can be made that most existing signal processing techniques have been proven appropriate for different ECG analysis tasks and it has been experimented with a large number of them. When it comes to machine learning algorithms, almost all of them are appropriate for ECG analysis tasks and have been examined. However, the overview of recent literature has shown that SVMs are one of the most commonly used overall at the moment, so it can be concluded that they are best-suitable for ECG analysis tasks.

From the overview in the paper, we can observe that deep neural networks have gained a lot of attention in recent years in ECG analysis. One of the most noteworthy feature of neural networks is that a lot of the ECG preprocessing steps covered so far are usually (not always, as in Labati et al. (2018)) omitted when working with them. The advantage of deep learning comes exactly from that: very little preprocessing and ECG expert knowledge is required when building the models. This is also one of its weaknesses: consequently, the resulting models are not explainable for the most part and probably for this reason they have not been so widely used in commercial products yet. The highest performance in some application areas (arrhythmia detection, biometric identification) was achieved recently using deep learning, which shows that these methods are a very promising direction for ECG analysis.

Most commercial software tools for ECG analysis are aimed at use in a clinical setting. The most widely-used example of this, the University of Glasgow software (Macfarlane et al., 2005), is only accurate for well-determined criteria for specific diagnoses. Overall, in these platforms, signal processing methods have successfully been implemented for some time by applying them to ECG signals for specific clinical tasks, usually related to ECG morphology, such as QRS detection. These software platforms, generally aimed at standard 12-lead ECG analysis or for Holter monitors, are usually implemented on the measurement device (Marsili et al., 2020), and still do not widely implement state-of-the-art deep learning techniques for end-to-end analysis. When it comes to machine learning and deep learning based diagnostic methods, they are most commonly realized as cloud-based platforms, such as Cardiologs (Smith, Walsh et al., 2019) or ViewECG.[22] As mentioned in the previous section, one of the challenges with mobile ECG sensors is the limited storage capacity and processing power, which brings new implementation challenges for all ECG analysis algorithms, including QRS detection (Elgendi et al., 2014). In this case, all analysis phases are generally performed at a remote location, either on a nearby processing device or as a remote service, with the device only transmitting the measurements. In some cases, after the measurement is over, only postponed offline analysis is performed, with a comprehensive analysis report obtained as a result.

4.3. Data and applications

In the second part, the paper has focused on the various data and application areas for ECG analysis. The goal of the critical overview of that part is to come to a conclusion which kinds of ECG data are already well-researched and which data-collection scenarios are still to be explored. In the overview of the databases, we can see that seemingly a large amount of ECG recordings is available. However, the largest portion of research uses only few datasets, most commonly the MIT-BIH Arrhythmia database, which is somehow understandable since arrhythmia detection is one of the most important areas of ECG analysis. Almost all of this data, however, is in a hospital setting, with a large portion being resting ECG. For example, the MIMIC dataset is very large and comprehensive, but only covers patients in the intensive care unit (ICU), which makes it suitable only for clinical tasks. With the exception of the Glasgow University Database, the iAmWell Dataset and the MARSH study, which are of relatively small size and the recordings are short, no extensive datasets consisting of recordings during daily movements and activities are available. Consequently, we can also conclude that no comprehensive databases with measurements from wireless body sensors are openly available. These kinds of data are necessary for all tasks related to daily monitoring of patients (monitoring of arrhythmia, physical exercise, breathing) and it is crucial that they are publicly available so that wider algorithm development and comparison are enabled.

When it comes to arrhythmia detection, there are a few databases (MIT-BIH Arrhythmia, INCART, MIT-BIH AF, AHA) containing labeled arrhythmia types and studies so far have achieved very high performance on them, most notably using recurrent and attention deep neural

[22] https://www.viewecg.com/.

networks. However, these databases contain a very limited number of examples for some arrhythmia types, and, as mentioned above, a limited variety of recording situations without including daily activities. The Stanford study (Hannun et al., 2019), in which a cardiologist-level arrhythmia detection and classification has been achieved using a deep neural network, is one example of a successful study in this aspect, but the data they collected has not been made publicly available yet. Furthermore, a novel 12-lead ECG database for arrhythmia research, covering more than 10,000 patients, is a very recent addition to this category of public ECG data (Zheng et al., 2020). Considering its large number of subjects and rhythm labels, it is a promising new direction for research in arrhythmia detection. Nevertheless, the research challenge set here is to utilize all 12 leads in the learning process for making more reliable models, hence more precise arrhythmia detection.

The use of databases in the task of biometric identification is very promising. Expensive labeling is not required for this task as it is for arrhythmia detection, so any database which includes some kind of patient ID could be used for this task. Nevertheless, the field is still lacking a thorough examination of the limits in regards to the number of subjects, that is, we need to know if the information that we can extract from the ECG is sufficient to distinguish a large population. One of the most recent additions to the PhysioNet repository – the PTB-XL dataset – could provide an answer to this question, because its previous version (PTB) was proven suitable for biometric identification, and could show whether deep learning is a better solution to this problem as well.

5. Conclusion and future challenges

The paper gives the most comprehensive survey of ECG databases and computational analysis methods and applications, including extensive overview of 45 public databases containing ECG recordings of various kinds and covering 7 ECG application areas: morphological and rhythmic arrhythmia detection, signal quality assessment, biometric identification, respiration estimation, fetal ECG extraction, and physical and emotional monitoring. The comprehensive survey in this paper provides an overview of what has been successfully achieved in the area of ECG analysis and what is just forming as a research direction. A general conclusion is that ECG for medical diagnosis is successfully analyzed with the existing methods, while different applications during daily ECG monitoring are still open fields.

Having this in mind, with the demonstrated multi-functionality of ECG monitoring, we can conclude that some kind of a general platform would be an ultimate goal of the ECG analysis pipeline. This platform would be flexible to different use-cases. Given how deep learning has been able to successfully address a lot of the most significant ECG problems, like arrhythmia classification, unification of different tasks under one platform could happen with deep learning, in the form of multi-task learning. In other ECG tasks, however, deep learning methods have not been comprehensively tested yet, such as the area of respiration and fetal ECG extraction. In addition, in order to be able to develop such an ECG analysis platform, data from a variety of settings needs to be available.

The lack of openly available comprehensive databases with ECG measurements from wireless body sensors indicates that one proposed future contribution is providing ECG data from mobile sensors to a public repository. In order to do this, a standardization of these measurements, recorded for a variety of applications, is needed first. Furthermore, a comprehensive evaluation of the best-performing methods for the applications covered in this paper, both on the measurements from mobile ECG devices, as well as the public databases, is also a possible direction for future work. Related to this, it is significant to find out how well the existing state-of-the-art methods for different tasks, trained on the open databases, perform when tested on new measurements, such as the ones from mobile ECG devices. This kind of standardized fair evaluation is an ultimate goal and will provide a more comprehensive analysis of exactly how much the challenges brought by mobile ECG sensors influence the performance of existing methods.

CRediT authorship contribution statement

Elena Merdjanovska: Methodology, Data curation, Visualization, Writing – original draft. **Aleksandra Rashkovska:** Conceptualization, Methodology, Supervision, Writing – original draft, 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.

Acknowledgments

The authors acknowledge the financial support of the Slovenian Research Agency under Grant P2-0095.

References

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., Gertych, A., & Tan, R. S. (2017). A deep convolutional neural network model to classify heartbeats. *Computers in Biology and Medicine*, *89*, 389–396. http://dx.doi.org/10.1016/j.compbiomed.2017.08.022.

Agrafioti, F., & Hatzinakos, D. (2010). Signal validation for cardiac biometrics. In *2010 IEEE international conference on acoustics, speech and signal processing* (pp. 1734–1737). http://dx.doi.org/10.1109/ICASSP.2010.5495461.

[dataset] Albrecht, P. (1992). The MIT-BIH ST change database. *PhysioNet*, http://dx.doi.org/10.13026/C2ZW2H.

Altini, M., Polito, S., Penders, J., Kim, H., Helleputte, N., Kim, S., & Yazicioglu, R. (2011). An ECG patch combining a customized ultra-low-power ECG SoC with bluetooth low energy for long term ambulatory monitoring. In *Proceedings of the 2nd conference on wireless health* (p. 15). http://dx.doi.org/10.1145/2077546.2077564.

An, X., & Stylios, G. (2020). Comparison of motion artefact reduction methods and the implementation of adaptive motion artefact reduction in wearable electrocardiogram monitoring. *Sensors*, *20*(1468), http://dx.doi.org/10.3390/s20051468.

Andreotti, F., Riedl, M., Himmelsbach, T., Wedekind, D., Wessel, N., Stepan, H., Schmieder, C., Jank, A., Malberg, H., & Zaunseder, S. (2014). Robust fetal ECG extraction and detection from abdominal leads. *Physiological Measurement*, *35*, 1551–1567. http://dx.doi.org/10.1088/0967-3334/35/8/1551.

ANSI/AAMI EC57 (2012). Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. In *Standard by the association for the advancement of medical instrumentation*.

Arzeno, N. M., Deng, Z.-D., & Poon, C.-S. (2008). Analysis of first-derivative based QRS detection algorithms. *IEEE Transactions on Biomedical Engineering*, *55*, 478–484. http://dx.doi.org/10.1109/TBME.2007.912658.

Baim, D. S., Colucci, W. S., Monrad, E. S., Smith, H. S., Wright, R. F., Lanoue, A., Gauthier, D. F., Ransil, B. J., Grossman, W., & Braunwald, E. (1986). Survival of patients with severe congestive heart failure treated with oral milrinone. *Journal of the American College of Cardiology*, *7*, 661–670. http://dx.doi.org/10.1016/S0735-1097(86)80478-8.

Barrett, K. E., Barman, S. M., Boitano, S., & Brooks, H. L. (2010). *Ganong's review of medical physiology* (Twenty-third ed.). New York: McGraw-Hill Medical.

Bassiouni, M., El-Dahshan, E.-S., Khalefa, W., & Salem, M. A.-B. (2018). Intelligent hybrid approaches for human ECG signals identification. *Signal, Image and Video Processing*, *12*, 941–949. http://dx.doi.org/10.1007/s11760-018-1237-5.

Behar, J., Bonnemains, L., Shulgin, V., Oster, J., Ostras, O., & Lakhno, I. (2018). Non-invasive fetal electrocardiography for the detection of fetal arrhythmias: Toward a fetal holter. *Archives of Cardiovascular Diseases Supplements*, *10*(281), http://dx.doi.org/10.1016/j.acvdsp.2018.06.019.

Behravan, V., Glover, N. E., Farry, R., Chiang, P. Y., & Shoaib, M. (2015). Rate-adaptive compressed-sensing and sparsity variance of biomedical signals. In *2015 IEEE 12th international conference on wearable and implantable body sensor networks (BSN)* (pp. 1–6). http://dx.doi.org/10.1109/BSN.2015.7299419.

Beraza, I., & Romero, I. (2017). Comparative study of algorithms for ECG segmentation. *Biomedical Signal Processing and Control*, *34*, 166–173. http://dx.doi.org/10.1016/j.bspc.2017.01.013.

Bousseljot, R.-D., Kreiseler, D., & Schnabel, A. (1995). Nutzung der EKG-signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik / Biomedical Engineering*, *40*, 317–318. http://dx.doi.org/10.1515/bmte.1995.40.s1.317.

Bramer, M. (2007). *Principles of data mining*. London: Springer, http://dx.doi.org/10.1007/978-1-84628-766-4.

Carreiras, C., Lourenco, A., Fred, A., & Ferreira, R. (2014). ECG signals for biometric applications - are we there yet? In *ICINCO 2014 - Proceedings of the 11th international conference on informatics in control, automation and robotics* (pp. 765–772). http://dx.doi.org/10.5220/0005160507650772.

Charlton, P. H., Birrenkott, D. A., Bonnici, T., Pimentel, M. A. F., Johnson, A. E. W., Alastruey, J., Tarassenko, L., Watkinson, P. J., Beale, R., & Clifton, D. A. (2018). Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review. *IEEE Reviews in Biomedical Engineering*, *11*, 2–20. http://dx.doi.org/10.1109/RBME.2017.2763681.

Charlton, P. H., Bonnici, T., Tarassenko, L., Clifton, D., Beale, R., & Watkinson, P. (2016). An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. *Physiological Measurement*, *37*, 610–626. http://dx.doi.org/10.1088/0967-3334/37/4/610.

Charlton, P. H., Villarroel, M., & Salguiero, F. (2016). Waveform analysis to estimate respiratory rate. In *Secondary analysis of electronic health records* (pp. 377–390). Cham: Springer, http://dx.doi.org/10.1007/978-3-319-43742-2_26.

Chauhan, S., & Vig, L. (2015). Anomaly detection in ECG time signals via deep long short-term memory networks. In *2015 IEEE international conference on data science and advanced analytics (DSAA)* (pp. 1–7). http://dx.doi.org/10.1109/DSAA.2015.7344872.

de Chazal, P., O'Dwyer, M., & Reilly, R. (2004). Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, *51*, 1196–1206. http://dx.doi.org/10.1109/TBME.2004.827359.

Chen, T.-M., Huang, C.-H., Shih, E. S., Hu, Y.-F., & Hwang, M.-J. (2020). Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model. *IScience*, *23*, Article 100886. http://dx.doi.org/10.1016/j.isci.2020.100886.

Chromik, J., Pirl, L., Beilharz, J., Arnrich, B., & Polze, A. (2021). Certainty in QRS detection with artificial neural networks. *Biomedical Signal Processing and Control*, *68*, Article 102628. http://dx.doi.org/10.1016/j.bspc.2021.102628.

Clifford, G., Liu, C., Moody, B., Lehman, L.-w., Silva, I., Li, Q., Johnson, A., & Mark, R. (2017). AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017. In *2017 computing in cardiology (CinC)* (pp. 1–4). http://dx.doi.org/10.22489/CinC.2017.065-469.

Clifford, G., Sameni, R., Ward, J., Robinson, J., & Wolfberg, A. J. (2011). Clinically accurate fetal ECG parameters acquired from maternal abdominal sensors. *American Journal of Obstetrics and Gynecology*, *205*(47), http://dx.doi.org/10.1016/j.ajog.2011.02.066, e1–47.e5.

Couderc, J. P., Xiaojuan, X., Zareba, W., & Moss, A. J. (2005). Assessment of the stability of the individual-based correction of qt interval for heart rate. In *Annals of noninvasive electrocardiology : the official journal of the international society for holter and noninvasive electrocardiology, inc* (pp. 25–34). http://dx.doi.org/10.1111/j.1542-474X.2005.00593.x, 10.

Da Poian, G., Bernardini, R., & Rinaldo, R. (2015). Separation and analysis of fetal-ECG signals from compressed sensed abdominal ECG recordings. *IEEE Transactions on Biomedical Engineering*, *63*, 1269–1279. http://dx.doi.org/10.1109/TBME.2015.2493726.

Demidova, M., Carlson, J., Erlinge, D., & Platonov, P. (2019). Early repolarization pattern on ECG recorded before the acute coronary event does not predict ventricular fibrillation during ST-elevation myocardial infarction. *Heart Rhythm*, *17*, 629–636. http://dx.doi.org/10.1016/j.hrthm.2019.11.011.

Depolli, M., Avbelj, V., Trobec, R., Kališnik, J. M., Korošec, T., Susič, A. P., Stanič, U., & Semeja, A. (2016). PCARD platform for mhealth monitoring. *Informatica*, *40*, 117–123.

Dinakarrao, S. M. P., Jantsch, A., & Shafique, M. (2019). Computer-aided arrhythmia diagnosis with bio-signal processing: A survey of trends and techniques. *ACM Computing Surveys*, *52*(23), http://dx.doi.org/10.1145/3297711.

Elgendi, M. (2013). Fast QRS detection with an optimized knowledge-based method: Evaluation on 11 standard ECG databases. *PLoS One*, *8*, Article e73557. http://dx.doi.org/10.1371/journal.pone.0073557.

Elgendi, M., Eskofier, B., Dokos, S., & Abbott, D. (2014). Revisiting QRS detection methodologies for portable, wearable, battery-operated, and wireless ECG systems. *PLoS One*, *9*, Article e84018. http://dx.doi.org/10.1371/journal.pone.0084018.

[dataset] Elia, R., Papa, K., Theocharides, T., & Orphanidou, C. (2017). *ECG, PPG, RSP signals from athletes and non-athletes during exercise - IAMwell*. Zenodo, http://dx.doi.org/10.5281/zenodo.1012726.

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. http://dx.doi.org/10.1016/j.cmpb.2018.04.005.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*, 861–874. http://dx.doi.org/10.1016/j.patrec.2005.10.010.

Fotiadou, E., Konopczyński, T., Hesser, J., & Vullings, R. (2020). End-to-end trained encoder–decoder convolutional neural network for fetal electrocardiogram signal denoising. *Physiological Measurement*, *41*, Article 015005. http://dx.doi.org/10.1088/1361-6579/ab69b9.

Gee, A. H., Barbieri, R., Paydarfar, D., & Indic, P. (2017). Predicting bradycardia in preterm infants using point process analysis of heart rate. *IEEE Transactions on Biomedical Engineering*, *64*, 2300–2308. http://dx.doi.org/10.1109/TBME.2016.2632746.

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P., Mark, R., Mietus, J., Moody, G., Peng, C.-K., & Stanley, H. (2000). PhysioBank, PhysioToolkit, and PhysioNet : Components of a new research resource for complex physiologic signals. *Circulation*, *101*, E215–20. http://dx.doi.org/10.1161/01.CIR.101.23.e215.

Greenwald, S. D. (1986). *Development and analysis of a ventricular fibrillation detector* (Master's thesis), MIT Dept. of Electrical Engineering and Computer Science.

Hamilton, P. (2002). Open source ECG analysis. In *Computers in cardiology* (pp. 101–104). http://dx.doi.org/10.1109/CIC.2002.1166717.

Hamilton, P. S., & Tompkins, W. J. (1986). Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. *IEEE Transactions on Biomedical Engineering, BME-33*, 1157–1165. http://dx.doi.org/10.1109/TBME.1986.325695.

Hannun, A., Rajpurkar, P., Haghpanahi, M., Tison, G., Bourn, C., Turakhia, M., & Ng, A. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, *25*, 65–69. http://dx.doi.org/10.1038/s41591-018-0268-3.

Healey, J., & Picard, R. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, *6*, 156–166. http://dx.doi.org/10.1109/TITS.2005.848368.

Hirsch, G., Jensen, S. H., Poulsen, E. S., & Puthusserypady, S. (2021). Atrial fibrillation detection using heart rate variability and atrial activity: A hybrid approach. *Expert Systems with Applications*, *169*, Article 114452. http://dx.doi.org/10.1016/j.eswa.2020.114452.

Hoefman, E., Bindels, P., & van Weert, H. (2010). Efficacy of diagnostic tools for detecting cardiac arrhythmias: systematic literature search. *NetherlandS Heart Journal : Monthly Journal of the NetherlandS Society of Cardiology and the Netherlands Heart Foundation*, *18*, 543–551. http://dx.doi.org/10.1007/s12471-010-0831-0.

Hong, S., Xiao, C., Ma, T., Li, H., & Sun, J. (2019). MINA: Multilevel knowledge-guided attention for modeling electrocardiography signals. In *Proceedings of the twenty-eighth international joint conference on artificial intelligence, IJCAI-19* (pp. 5888–5894). http://dx.doi.org/10.24963/ijcai.2019/816.

[dataset] Howell, L., & Porr, B. (2018). *High precision ECG database with annotated R peaks, recorded and filmed under realistic conditions*. University of Glasgow, http://dx.doi.org/10.5525/gla.researchdata.716.

Huang, H., Hu, S., & Sun, Y. (2019). A discrete curvature estimation based low-distortion adaptive Savitzky–Golay filter for ECG denoising. *Sensors*, *19*(1617), http://dx.doi.org/10.3390/s19071617.

IEC 60601-2-47 (2012). Medical electrical equipment - part 2-47: Particular requirements for the basic safety and essential performance of ambulatory electrocardiographic systems. In *Standard by the international electrotechnical commission*.

Ilic, V., Trobec, R., & Širaiy, B. (2019). Quality of one-channel telemetric ECG sensor signal in maximum exercise stress tests. *Measurement Science Review*, *19*, 79–85. http://dx.doi.org/10.2478/msr-2019-0013.

Iqbal, T., Elahi, A., Redon, P., Vazquez, P., Wijns, W., & Shahzad, A. (2021). A review of biophysiological and biochemical indicators of stress for connected and preventive healthcare. *Diagnostics*, *11*(556), http://dx.doi.org/10.3390/diagnostics11030556.

Iyengar, N., Peng, C.-K., Morin, R., Goldberger, A., & Lipsitz, L. (1996). Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *The American Journal of Physiology*, *271*, R1078–R1084. http://dx.doi.org/10.1152/ajpregu.1996.271.4.R1078.

Jezewski, J., Matonia, A., Kupka, T., Roj, D., & Czabanski, R. (2012). Determination of fetal heart rate from abdominal signals: Evaluation of beat-to-beat accuracy in relation to the direct fetal electrocardiogram. *Biomedizinische Technik/Biomedical Engineering*, *57*, 383–394. http://dx.doi.org/10.1515/bmt-2011-0130.

Kalyakulina, A. I., Yusipov, I. I., Moskalenko, V. A., Nikolskiy, A. V., Kosonogov, K. A., Osipov, G. V., Zolotykh, N. Y., & Ivanchenko, M. V. (2020). LUDB: A new open-access validation tool for electrocardiogram delineation algorithms. *IEEE Access*, *8*, 186181–186190. http://dx.doi.org/10.1109/ACCESS.2020.3029211.

Khawaja, A., Litwin, J., Auzinger, T., O'Rourke, W., Devine, T., Furlong, A., Lehmann, C., & Fischer, R. (2011). Analyzing the delineation precision of hannover ECG system (HES®): A validation study. In *2011 computing in cardiology* (pp. 617–620).

[dataset] Kher, R. (2020). Wearable ambulatory electrocardiogram (ECG) and EEG dataset. *IEEE DataPort*, http://dx.doi.org/10.21227/YSNC-GC65.

Kiranyaz, S., Ince, T., & Gabbouj, M. (2017). Personalized monitoring and advance warning system for cardiac arrhythmias. *Scientific Reports*, *7*, 9270. http://dx.doi.org/10.1038/s41598-017-09544-z.

Labati, R., Muñoz, E., Piuri, V., Sassi, R., & Scotti, F. (2018). Deep-ECG: Convolutional neural networks for ECG biometric recognition. *Pattern Recognition Letters*, *126*, 78–85. http://dx.doi.org/10.1016/j.patrec.2018.03.028.

Laguna, P., Mark, R., Goldberg, A., & Moody, G. (1997). Database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In *Computers in cardiology 1997* (pp. 673–676). http://dx.doi.org/10.1109/CIC.1997.648140.

Li, Y., Pang, Y., Wang, K., & Li, X. (2020). Toward improving ECG biometric identification using cascaded convolutional neural networks. *Neurocomputing*, *391*, 83–95. http://dx.doi.org/10.1016/j.neucom.2020.01.019.

Li, H., Yuan, D., Ma, X., Cui, D., & Cao, L. (2017). Genetic algorithm for the optimization of features and neural networks in ECG signals classification. *Scientific Reports*, *7*(41011), http://dx.doi.org/10.1038/srep41011.

Liu, F., Liu, C., Zhao, L., Zhang, X., Wu, X., Xu, X., Liu, Y., Ma, C., Wei, S., He, Z., Li, J., & Ng, E. (2018). An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *Journal of Medical Imaging and Health Informatics*, *8*, 1368–1373. http://dx.doi.org/10.1166/jmihi.2018.2442.

[dataset] Lugovaya, T. (2011). The ECG-ID database. *PhysioNet*, http://dx.doi.org/10.13026/C2J01F.

Luz, E., Schwartz, W., Chávez, G., & Menotti, D. (2015). ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer Methods and Programs in Biomedicine*, *127*, 144–164. http://dx.doi.org/10.1016/j.cmpb.2015.12.008.

Lyon, A., Mincholé, J. P., Laguna, P., & Rodriguez, B. (2018). Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. *Journal of the Royal Society Interface*, *15*, Article 20170821. http://dx.doi.org/10.1098/rsif.2017.0821.

Macfarlane, P., Browne, D., Devine, B., Clark, E., Miller, E., Seyal, J., & Hampton, D. (2004). Modification of ACC/ESC criteria for acute myocardial infarction. *Journal of Electrocardiology*, *37 Suppl*, 98–103. http://dx.doi.org/10.1016/j.jelectrocard.2004.08.032.

Macfarlane, P., Devine, B., & Clark, E. (2005). The university of glasgow (Uni-G) ECG analysis program. In *Computers in cardiology* (pp. 451–454). http://dx.doi.org/10.1109/CIC.2005.1588134.

Macfarlane, P. W., Oosterom, A. Van., Pahlm, O., Kligfield, P., Janse, M., & Camm, J. (Eds.), (2010). *Comprehensive electrocardiology* (2nd ed.). London: Springer, http://dx.doi.org/10.1007/978-1-84882-046-3.

Markova, V., Ganchev, T., & Kalinkov, K. (2019). CLAS: A database for cognitive load, affect and stress recognition. In *2019 International conference on biomedical innovations and applications (BIA)* (pp. 1–4). http://dx.doi.org/10.1109/BIA48344.2019.8967457.

Marsili, I. A., Biasiolli, L., Masè, M., Adami, A., Andrighetti, A. O., Ravelli, F., & Nollo, G. (2020). Implementation and validation of real-time algorithms for atrial fibrillation detection on a wearable ECG device. *Computers in Biology and Medicine*, *116*, Article 103540. http://dx.doi.org/10.1016/j.compbiomed.2019.103540.

Matonia, A., Jezewski, J., Kupka, T., Jezewski, M., Horoba, K., Wrobel, J., Czabanski, R., & Kahankova, R. (2020). Fetal electrocardiograms, direct and abdominal with reference heartbeat annotations. *Scientific Data*, *7*, 200. http://dx.doi.org/10.1038/s41597-020-0538-z.

Mei, Z., Gu, X., Chen, H., & Chen, W. (2018). Automatic atrial fibrillation detection based on heart rate variability and spectral features. *IEEE Access*, *6*, 53566–53575. http://dx.doi.org/10.1109/ACCESS.2018.2871220.

[dataset] Moody, G. (1999). *The MIT-BIH normal sinus rhythm database*. PhysioNet, http://dx.doi.org/10.13026/C2NK5R.

Moody, G., & Mark, R. (1983). A new method for detecting atrial fibrillation using R-R intervals. *Computers in Cardiology*, *10*, 227–230. http://dx.doi.org/10.1093/europace/eum096.

Moody, G., & Mark, R. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, *20*, 45–50. http://dx.doi.org/10.1109/51.932724.

Moody, G., Muldrow, W., & Mark, R. (1984). A noise stress test for arrhythmia detectors. *Computers in Cardiology*, *11*, 381–384.

Mousavi, S., & Afghah, F. (2019). Inter- and intra- patient ECG heartbeat classification for arrhythmia detection: A sequence to sequence deep learning approach. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 1308–1312). http://dx.doi.org/10.1109/ICASSP.2019.8683140.

Mousavi, S., Afghah, F., Razi, A., & Acharya, U. R. (2019). ECGNET: Learning where to attend for detection of atrial fibrillation with deep visual attention. In *2019 IEEE EMBS international conference on biomedical health informatics (BHI)* (pp. 1–4). http://dx.doi.org/10.1109/BHI.2019.8834637.

Nolle, F. M., Badura, F. K., Catlett, J. M., Bowser, R. W., & Sketch, M. H. (1986). CREI-GARD, a new concept in computerized arrhythmia monitoring systems. *Computers in Cardiology*, *13*, 515–518.

Oweis, R. J., & Al-Tabbaa, B. O. (2014). QRS detection and heart rate variability analysis: A survey. *Biomedical Science and Engineering*, *2*, 13–34.

Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, *BME-32*, 230–236. http://dx.doi.org/10.1109/TBME.1985.325532.

Petrutiu, S., Sahakian, A., & Swiryn, S. (2007). Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. *Europace*, *9*, 466–470. http://dx.doi.org/10.1093/europace/eum096.

Pimentel, M. A. F., Charlton, P. H., & Clifton, D. A. (2015). Probabilistic estimation of respiratory rate from wearable sensor. In S. Mukhopadhyay (Ed.), *Wearable electronics sensors* (pp. 241–262). Cham: Springer, http://dx.doi.org/10.1007/978-3-319-18191-2_10.

Pirhonen, M., & Vehkaoja, A. (2020a). Fusion enhancement for tracking of respiratory rate through intrinsic mode functions in photoplethysmography. *Biomedical Signal Processing and Control*, *59*, Article 101887. http://dx.doi.org/10.1016/j.bspc.2020.101887.

[dataset] Pirhonen, M., & Vehkaoja, A. (2020b). *'MARSH' - Respiratory signal repository*. Zenodo, http://dx.doi.org/10.5281/ZENODO.3673923.

Porr, B., & Howell, L. (2019). R-peak detector stress test with a new noisy ECG database reveals significant performance differences amongst popular detectors. *BioRxiv*, Article 722397. http://dx.doi.org/10.1101/722397.

Qin, Q., Li, J., Zhang, L., Yue, Y., & Liu, C. (2017). Combining low-dimensional wavelet features and support vector machine for arrhythmia beat classification. *Scientific Reports*, *7*, 6067. http://dx.doi.org/10.1038/s41598-017-06596-z.

Raj, S., & Ray, K. (2018). A personalized arrhythmia monitoring platform. *Scientific Reports*, *8*, 11395. http://dx.doi.org/10.1038/s41598-018-29690-2.

Rashkovska, A., & Avbelj, V. (2017). Abdominal fetal ECG measured with differential ECG sensor. In *2017 40th international convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 289–291). http://dx.doi.org/10.23919/MIPRO.2017.7973436.

Rashkovska, A., Depolli, M., Tomašić, I., Avbelj, V., & Trobec, R. (2020). Medical-grade ECG sensor for long-term monitoring. *Sensors*, *20*, 1695. http://dx.doi.org/10.3390/s20061695.

Ribeiro, A. H., Ribeiro, M., Paixão, G., Oliveira, D., Gomes, P., Canazart, J., Ferreira, M., Andersson, C., Macfarlane, P., Wagner, M., Schön, T., & Ribeiro, A. L. (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature Communications*, *11*, 1760. http://dx.doi.org/10.1038/s41467-020-15432-4.

Salloum, R., & Kuo, C.-C. J. (2017). ECG-based biometrics using recurrent neural networks. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 2062–2066). http://dx.doi.org/10.1109/ICASSP.2017.7952519.

Sameni, R., Shamsollahi, M. B., Jutten, C., & Clifford, G. D. (2007). A nonlinear Bayesian filtering framework for ECG denoising. *IEEE Transactions on Biomedical Engineering*, *54*, 2172–2185. http://dx.doi.org/10.1109/TBME.2007.897817.

Satija, U., Ramkumar, B., & Manikandan, M. S. (2019). A new automated signal quality-aware ECG beat classification method for unsupervised ECG diagnosis environments. *IEEE Sensors Journal*, *19*, 277–286. http://dx.doi.org/10.1109/JSEN.2018.2877055.

Sayadi, O. (2008). ECG denoising and compression using a modified extended Kalman filter structure. *IEEE Transactions on Biomedical Engineering*, *55*, 2240–2248. http://dx.doi.org/10.1109/TBME.2008.921150.

Sellami, A., & Hwang, H. (2018). A robust deep convolutional neural network with batch-weighted loss for heartbeat classification. *Expert Systems with Applications*, *122*, 75–84. http://dx.doi.org/10.1016/j.eswa.2018.12.037.

Seshadri, D., Li, R., Voos, J., Rowbottom, J., Alfes, C., Zorman, C., & Drummond, C. (2019). Wearable sensors for monitoring the internal and external workload of the athlete. *npj Digital Medicine*, *2*, 71. http://dx.doi.org/10.1038/s41746-019-0149-2.

Sharma, K., Castellini, C., van den Broek, E. L., Albu-Schäeffer, A., & Schwenker, F. (2019). A dataset of continuous affect annotations and physiological signals for emotion analysis. *Scientific Data*, *6*, 196. http://dx.doi.org/10.1038/s41597-019-0209-0.

Silva, I., Behar, J., Sameni, R., Zhu, T., Oster, J., Clifford, G. D., & Moody, G. B. (2013). Noninvasive fetal ECG: the PhysioNet/computing in cardiology challenge 2013. In *Computing in cardiology 2013* (pp. 149–152).

Smith, S., Rapin, J., Li, J., Fleureau, Y., Fennell, W., Walsh, B., Rosier, A., Fiorina, L., & Gardella, C. (2019). A deep neural network for 12-lead electrocardiogram interpretation outperforms a conventional algorithm, and its physician overread, in the diagnosis of atrial fibrillation. *IJC Heart & Vasculature*, *25*, Article 100423. http://dx.doi.org/10.1016/j.ijcha.2019.100423.

Smith, S. W., Walsh, B., Grauer, K., Wang, K., Rapin, J., Li, J., Fennell, W., & Taboulet, P. (2019). A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. *Journal of Electrocardiology*, *52*, 88–95. http://dx.doi.org/10.1016/j.jelectrocard.2018.11.013.

Sohn, K., Merchant, F., Sayadi, O., Puppala, D., Doddamani, R., Sahani, A., Singh, J., Heist, E., Isselbacher, E., & Armoundas, A. (2017). A novel point-of-care smartphone based system for monitoring the cardiac and respiratory systems. *Scientific Reports*, *7*, 44946. http://dx.doi.org/10.1038/srep44946.

Spetlik, R., Franc, V., Cech, J., & Matas, J. (2018). Visual heart rate estimation with convolutional neural network. In *Proceedings of the British machine vision conference*.

Šprager, S., Trobec, R., & Juric, M. (2017). Feasibility of biometric authentication using wearable ECG body sensor based on higher-order statistics. In *2017 40th international convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 264–269). http://dx.doi.org/10.23919/MIPRO.2017.7973431.

Strodthoff, N., Wagner, P., Schaeffter, T., & Samek, W. (2021). Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *IEEE Journal of Biomedical and Health Informatics*, *25*, 1519–1528. http://dx.doi.org/10.1109/JBHI.2020.3022989.

[dataset] Tihonenko, V., Khaustov, A., Ivanov, S., & Rivin, A. (2007). *St.-Petersburg institute of cardiological technics 12-lead arrhythmia database*. PhysioNet, http://dx.doi.org/10.13026/C2V88N.

Tracey, B. H., & Miller, E. L. (2012). Nonlocal means denoising of ECG signals. *IEEE Transactions on Biomedical Engineering*, *59*, 2383–2386. http://dx.doi.org/10.1109/TBME.2012.2208964.

Trobec, R., Avbelj, V., & Rashkovska, A. (2014). Multi-functionality of wireless body sensors. *The IPSI BgD Transactions on Internet Research*, *10*, 23–27.

Trobec, R., Rashkovska, A., & Avbelj, V. (2012). Two proximal skin electrodes — A respiration rate body sensor. *Sensors*, *12*, 13813–13828. http://dx.doi.org/10.3390/s121013813.

Trobec, R., Tomašić, I., Rashkovska, A., Depolli, M., & Avbelj, V. (2018). *Body sensors and electrocardiography*. Springer, http://dx.doi.org/10.1007/978-3-319-59340-1.

Turakhia, M., Hoang, D., Zimetbaum, P., Miller, J., Froelicher, V., Kumar, U., Xu, X., Yang, F., & Heidenreich, P. (2013). Diagnostic utility of a novel leadless arrhythmia monitoring device. *The American Journal of Cardiology*, *112*, 520–524. http://dx.doi.org/10.1016/j.amjcard.2013.04.017.

Varon, C., Morales, J., Lazaro, J., Orini, M., Deviaene, M., Kontaxis, S., Testelmans, D., Buyse, B., Borzée, P., Sörnmo, L., Laguna, P., Gil, E., & Bailón, R. (2020). A comparative study of ECG-derived respiration in ambulatory monitoring using the single-lead ECG. *Scientific Reports*, *10*, 5704. http://dx.doi.org/10.1038/s41598-020-62624-5.

Wagner, P., Strodthoff, N., Bousseljot, R., Kreiseler, D., Lunze, F., Samek, W., & Schaeffter, T. (2020). PTB-XL: A large publicly available electrocardiography dataset. *Scientific Data*, *7*, 154. http://dx.doi.org/10.1038/s41597-020-0495-6.

Weng, B., Blanco-Velasco, M., & Barner, K. E. (2006). ECG denoising based on the empirical mode decomposition. In *2006 international conference of the IEEE engineering in medicine and biology society* (pp. 1–4). http://dx.doi.org/10.1109/IEMBS.2006.259340.

WHO (2021). *Cardiovascular diseases (CVDs)*: *Fact sheet*, [Online]. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

Xiang, Y., Lin, Z., & Meng, J. (2018). Automatic QRS complex detection using two-level convolutional neural network. *BioMedical Engineering OnLine*, *17*, 13. http://dx.doi.org/10.1186/s12938-018-0441-4.

Yao, Q., Wang, R., Fan, X., Liu, J., & Li, Y. (2020). Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network. *Information Fusion*, *53*, 174–182. http://dx.doi.org/10.1016/j.inffus.2019.06.024.

Yildirim, O., Talo, M., Ciaccio, E. J., Tan, R. S., & Acharya, U. R. (2020). Accurate deep neural network model to detect cardiac arrhythmia on more than 10, 000 individual subject ECG records. *Computer Methods and Programs in Biomedicine*, *197*, Article 105740. http://dx.doi.org/10.1016/j.cmpb.2020.105740.

Zhai, X., Zhoua, Z., & Tin, C. (2020). Semi-supervised learning for ECG classification without patient-specific labeled data. *Expert Systems with Applications*, *158*, Article 113411. http://dx.doi.org/10.1016/j.eswa.2020.113411.

Zhang, T., El Ali, A., Wang, C., Hanjalic, A., & Cesar, P. (2021). CorrNet: Fine-grained emotion recognition for video watching using wearable physiological sensors. *Sensors*, *21*, 52. http://dx.doi.org/10.3390/s21010052.

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*, 48. http://dx.doi.org/10.1038/s41597-020-0386-x.

Zywietz, C., Borovsky, D., Goetsch, G., & Joseph, G. (1990). Methodology of ECG interpretation in the Hannover program. *Methods of Information in Medicine*, *29*, 375–385. http://dx.doi.org/10.1055/s-0038-1634800.