Journal of Electrocardiology 57 (2019) S61–S64

Contents lists available at ScienceDirect

Journal of Electrocardiology

journal homepage: www.jecgonline.com

Review

Machine learning in the electrocardiogram

Ana Mincholé, PhD [a,*], Julià Camps, MSc [a], Aurore Lyon, PhD [b], Blanca Rodríguez, PhD [a]

[a] Department of Computer Science, University of Oxford, Oxford, United Kingdom
[b] CARIM, Maastricht University, Maastricht, Netherlands

ARTICLE INFO

Article history:
Received 13 June 2019
revised 30 July 2019
accepted 8 August 2019

ABSTRACT

The electrocardiogram is the most widely used diagnostic tool that records the electrical activity of the heart and, therefore, its use for identifying markers for early diagnosis and detection is of paramount importance. In the last years, the huge increase of electronic health records containing a systematised collection of different type of digitalised medical data, together with new tools to analyse this large amount of data in an efficient way have re-emerged the field of machine learning in healthcare innovation. This review describes the most recent machine learning-based systems applied to the electrocardiogram as well as pros and cons in the use of these techniques. Machine learning, including deep learning, have shown to be powerful tools for aiding clinicians in patient screening and risk stratification tasks. However, they do not provide the physiological basis of classification outcomes. Computational modelling and simulation can help in the interpretation and understanding of key physiologically meaningful ECG biomarkers extracted from machine learning techniques.

Introduction

Cardiovascular diseases are the main cause of death in industrialised countries accounting for 17.9 million deaths each year, 31% of all deaths worldwide [1]. Consequently, identifying markers for early diagnosis and detection, and timely treatment is of paramount importance. The widely known electrocardiographic (ECG) signal is one of the most used clinical tools for the evaluation of the heart function with the advantage of being non-invasive and non-expensive. In the clinical practice, the ECG signal is usually interpreted by an electrophysiologist who requires a high level of expertise. ECG interpretation is time-consuming and highly dependent on individual interpretation. In the new technological healthcare era, the scenario is moving towards an effective quantification and analysis of the ECG signal that assists clinicians in the evaluation of patients' risk.

In the last years, the huge increase of electronic health records containing systematised collections of different types of digitalised medical data, together with new tools to analyse this large amount of data in an efficient way have re-emerged the field of machine learning (ML) in healthcare innovation [2]. Furthermore, hardware improvements over the last decade, such as the availability of powerful computing platforms as cloud computing, high-performance computing or graphics processing units (GPUs), together with efficient modern software techniques have supported the development of new technologies in ML, and specifically deep learning [3,4]. These rely entirely on the data and automatically identify data and characteristic patterns for a specific task [4].

These new technologies have provided great success in fields such as speech recognition [3], or image analysis [5]. In the area of clinical electrophysiology, the storage and curation of large digitised datasets in some hospitals together with new computing platforms have led to the development and adoption of new ML technologies to maximise the information extracted from comprehensive ECG datasets [6,7].

In this review, we describe new advances in the ECG analysis for the identification and classification of cardiovascular diseases and conditions, and, the pros and cons of different ML techniques, and their synergy with computational modelling and simulation.

Supervised versus unsupervised learning

Within the ML framework, there are two main types of tasks: supervised and unsupervised [8]. Most of the ML studies performed in the area of ECG analysis aimed at disease diagnosis or risk stratification use supervised learning techniques. The goal of supervised learning is the inference of a function or a score from labelled or annotated training data. The two main supervised techniques are classification and regression, which differ on having as outputs categorical and continuous variables, respectively.

Classification of different heart rhythms is one of the most developed applications of ML to the ECG using supervised approaches (see for example [6,9,10]). The main reason is the availability of public databases such as MIT-BIH Arrhythmia [11] from Physionet Project (https://physionet.org/physiobank/database/) which provides large datasets to train the algorithms. Some popular algorithms include logistic

* Corresponding author at: Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD, United Kingdom.
E-mail address: ana.minchole@gmail.com (A. Mincholé).

https://doi.org/10.1016/j.jelectrocard.2019.08.008
0022-0736/

regression, support vector machines, artificial neural networks, and random forests as reviewed in [9].

On the other hand, unsupervised learning aims at discovering *hidden* structures in datasets with no previous knowledge about reference outcomes or labels [8]. The most common unsupervised learning method is clustering, used for grouping data with a similar data structure. Another common task in unsupervised learning is dimensionality reduction where principal component analysis (PCA) is one of the most frequently used methods in traditional ECG analysis, projecting data onto its feature subspace [12]. A recent case of success has been the use of mathematical modelling based on the Hermite functions and clustering techniques to identify ECG phenotypic subgroups in hypertrophic cardiomyopathy [13]. The identified ECG phenotypes (Fig. 1B) or clusters were based on morphological features of the QRS complex and on T wave polarity [13]. These clusters were subsequently associated with arrhythmic risk markers.

Traditional ML techniques versus deep learning

In this review, we refer to traditional ML techniques as the ones that require the initial definition of a feature vector to describe the ECG beat. Feature extraction is followed by training the classifier over the extracted features. Most of the studies on the classification of ECG recordings or beats rely on the derivation of ECG based features [9,14]. Features based on time intervals such as QRS duration, QT interval, heart rate, or morphological features based on wavelet or Hermite coefficients have been widely used because of their clinical interpretability [9]. The limited amount of data in many databases with respect to the number of potential features describing them requires the use of techniques as feature extraction and dimensionality reduction to avoid overfitting and, in turn, the lack of generalisation to other databases. Overfitting is a common problem in ML algorithms that occurs when the model fits well to the training set by modelling its particularities but does not perform well when evaluated in an unseen new dataset or test set. Additionally, traditional ML techniques usually allow the identification of the most significant features and their contribution to the final score providing interpretability to the ML system [14].

As opposed to ML models using *hand-crafted* features, artificial neural networks (ANNs) are designed to automatically extract the optimal features for performing a specific task with the drawback of being hard to interpret.

The simplest neural network is the perceptron, which works as a type of artificial neuron. It consists of a transfer function (a weighted sum of the inputs) followed by a non-linear activation function such as a sigmoid function or a rectifier linear unit (ReLU) resulting in the output of the neuron.

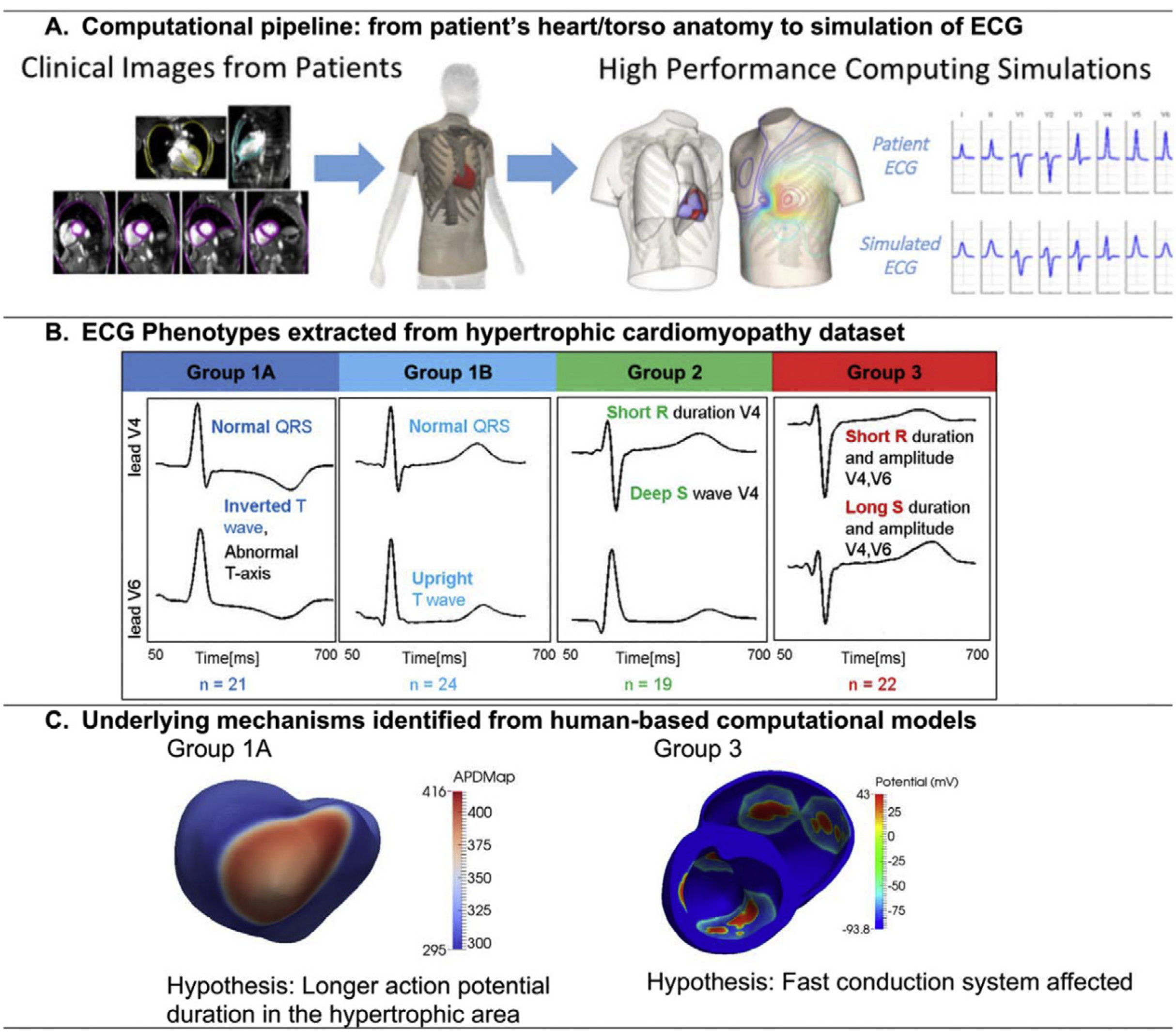

Fig. 1. Computational modelling and simulation to understand the rationale behind the ML outcomes. Panel A shows the computational pipeline from patient specific heart/torso geometry to the simulation of the ECG through biophysically detailed computational models (image credits from [19]). Panel B shows the four ECG phenotypes extracted from clustering techniques (image credits from [13]). Panel C shows the mechanisms identified by computer models that explain the ECG phenotypes (image credits from [20]).

Neurons can be arranged into interconnected layers, the first being one the input layer and the last one the output layer. Once the architecture of an ANN is defined, the training process consists of optimising its performance by adjusting the weights of the connections between artificial neurons so that the deep learning model learns what combinations of patterns explains best the data [4].

Architectures with many layers are considered deep neural networks (DNNs) [4], among which the most popular ones are the fully connected neural networks (FNNs) and the convolutional neural networks (CNNs) [4]. The FNNs are networks with multiple layers with each neuron connected to every neuron of the previous layer and each connection having its own weight. On the other hand, CNNs take advantage of the hierarchical pattern in data and assemble more complex patterns and at different scales, improving filter bank approaches [10]. A CNN is then able to successfully capture patterns in the 1D ECG signal through the application of relevant filters. Each layer includes a non-linear convolutional layer for extracting the features and performing the activation followed by a pool layer, which performs a sub-sampling on the output of the convolutional layer. The initial convolutional layers exploit the local information or low-scale patterns of the signal while the last ones learn more global and higher-level patterns of the input data relevant to perform specific tasks.

The time series nature of the electrocardiogram makes it a good candidate to use recurrent neural networks able to effectively capture the temporal characteristics of the signal. Therefore, some works as for example [15], focused on ECG classification have successfully combined convolutional layers for feature extraction with long-short term memory (LSTM) layers, a recurrent neural network architecture, allowing temporal aggregation of features.

Deep learning techniques benefit from large quantities of data and from optimised methodologies to efficiently deal with these data in reasonable amounts of time. However, the number of parameters/connections makes these architectures prone to overfitting data. The most common strategies to reduce overfitting are through early termination methods, DropOut and regularized cost functions [4].

The different ML approaches present advantages and disadvantages. In the context of ECG classification, the performance of ML approaches such as random forests and linear techniques strongly depends on the characterisation power of the features extracted from the ECG data. However, since the contribution of each extracted feature in the classification output is known, results from traditional ML are typically clinically interpretable as reviewed in [9].

On the other hand, ANNs including deep learning approaches learn the optimal features and parameters for a specific task through the use of data. These optimal features are typically difficult to interpret or to provide them with a physiological explanation from the clinical knowledge, which has always relied on pre-defined biomarkers. This loss of interpretability gives the feeling of dealing with a “black box” as it is not possible to make assumptions about functional dependencies between inputs and outputs.

Deep learning and its use for patient classification

Two recent papers [6,7] demonstrate the power of deep learning applied to the electrocardiogram benefitting from large quantities of data (of over 90,000 ECG recordings for training the models) and from optimised methodologies to deal with these data efficiently. Both present DNNs in an end-to-end manner.

The first study [6] aimed at detecting different types of arrhythmia by classifying 12 rhythm classes such as atrial fibrillation, atrioventricular block, ventricular tachycardia or sinus rhythm as well as noisy recordings. They trained a 34-layer DNN with over 90,000 single lead ambulatory ECG and validated the performance against an independent test dataset of 300 recordings annotated by a certified board of cardiologists. Results show an area under the curve (AUC) over 0.91 for each the individual rhythm classes. AUC is a measure of the performance resulting in 1.0 for a perfect classification and 0.5 for a completely random classification.

Even if most of the previous ECG classification studies described in [9,10], present comparable accuracies over 90%, the main disadvantage with respect to the recent [6], is the smaller datasets used. This may limit the generalisation properties when tested on new databases.

The second study [7] presented a screening system for the identification of patients with left ventricular systolic dysfunction (ejection fraction <35%) from 12-lead ECGs, using a 6-layer CNN. This CNN was trained by using paired 12-lead ECGs and echocardiogram from around 45,000 patients and tested on >52,000 patients. The obtained AUC curve was 0.93 improving the current B-type natriuretic peptide (BNP) screening blood tests with an AUC of 0.79–0.89.

Deep learning and its use for pre-processing tasks

As an illustration of a successful study with effects in the specific area of cardiology, medical image analysis has experienced an absolute renovation through ML showing great potential in image segmentation and analysis. The use of deep learning has allowed to automatically perform repetitive and daunting tasks, mainly for a pre-processing stage, such as segmentation of heart chambers in cardiac magnetic resonance (MR) images. A recent study [5] has demonstrated a human-level performance of a fully CNN for MR image segmentation, being a starting point for automated MR analysis on a large database. On the contrary, its counterpart in the ECG signal, which is the delineation of the ECG waveforms, remains a challenge.

The analysis of ECG recordings in the clinical practice usually involves computing standard biomarkers such as heart rate, QRS width, QT interval and ST segment elevation. These biomarkers are fundamental for clinicians and automatic decision-making systems to determine specific cardiac condition and to monitor events such as ventricular arrhythmias. Their computation requires a first step consisting of the delineation of the different ECG waveforms by identifying fiducial points such as onset and offset of the QRS complex, R wave peak, and T-wave onset, peak and offset.

The state-of-the-art ECG delineation systems [16] show good performance when tested on the available and scarce annotated databases that are small lacking from variability in their data. As an example, the QT database [17] (QTDB) considered only recordings with low levels of noise and therefore most of the recordings present slow heart rates. Therefore, most of the automatic delineation tools lead to poor generalisation when using different databases than the ones considered in the original design.

Deep learning algorithms are designed for learning the data structure automatically without the need of adding expert knowledge into the model. Additionally, ML techniques such as autoencoders [4] allow the use of unlabelled or non-annotated ECG data in the design of the delineation system improving generalisation by the use of increased variability in the training stage.

Recently, a deep learning-based multilead ECG delineation method was presented [18], showing performances in delineating the QRS complex comparable to the state-of-the-art when evaluated on the QTDB.

Human-based computer modelling and simulation for the interpretability of ML outcomes

Whereas ML can provide metrics for risk stratification and classification, it is sometimes difficult to interpret the functional dependencies between outputs and inputs, as well as the importance of each feature on the model predictions, and how the different features interact. This lack of interpretability in ML models can potentially have adverse or even life-threatening consequences, hindering the acceptance by healthcare communities. In the case of knowing or having a clinically interpretable physiological model, it is still technically difficult to include it in the ML classifier, and more in particular, when using deep learning approaches.

Human biophysically detailed computational modelling and simulation may be used to aid in the interpretation of findings from ML studies by linking structural and electrophysiological changes to ECG abnormalities [19] (see Fig. 1A). As a successful example of its potential, the study described in [20] aimed at the identification of the underlying mechanisms explaining distinct ECG ML-based phenotypes [13] in hypertrophic cardiomyopathy (see Fig. 1B). The ECG phenotype with T wave inversion was explained by action potential prolongation in hypertrophied area, whereas abnormal QRS complex was explained by a poor Purkinje-myocardial coupling (see Fig. 1C).

Conclusions

ML algorithms applied to the electrocardiogram are powerful tools for assisting in the tasks of patient screening and have the potential to advance patient stratification by taking advantage of population-based information. Recent DNN approaches have already shown a successful performance in identifying abnormal heart rhythms and mechanical dysfunction. DNNs efficiently process a large amount of data, being ideal for real-time and wearable devices in need of reliable and efficient analysis of ECG signals.

On the other hand, understanding the rationale behind their results becomes a challenge, as ML methods are unable to reveal the functional dependencies between inputs and outputs. ML can be used in combination with other techniques such as computational modelling and simulation to explain the rationale behind the ML outcomes.

In general, ML techniques rely on data and the variability, quality and magnitude of signals in the training dataset impose limits to the model performance. The model can be as good as the relevant information in the dataset.

Acknowledgements

AM and BR are supported by a Wellcome Trust Senior Research Fellowship in Basic Biomedical Sciences and the British Heart Foundation Centre of Research Excellence. JC is supported by The Engineering and Physical Sciences Research Council. This project also received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 675451 (CompBioMed project).

References

[1] WHO | Cardiovascular diseases (CVDs). WHO n.d. http://www.who.int/mediacentre/factsheets/fs317/en/ (accessed February 3, 2018).

[2] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203. https://doi.org/10.1038/s41586-018-0579-z.

[3] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44. https://doi.org/10.1038/nature14539.

[4] Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016.

[5] Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. J Cardiovasc Magn Reson. 2018;20:65. https://doi.org/10.1186/s12968-018-0471-x.

[6] Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med. 2019;25:65. https://doi.org/10.1038/s41591-018-0268-3.

[7] Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nat Med. 2019;25:70. https://doi.org/10.1038/s41591-018-0240-2.

[8] Bishop C. Pattern recognition and machine learning. New York: Springer-Verlag; 2006.

[9] Lyon A, Mincholé A, Martínez JP, Laguna P, Rodriguez B. Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. J R Soc Interface. 2018;15. https://doi.org/10.1098/rsif.2017.0821.

[10] Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans Biomed Eng. 2016;63:664–75. https://doi.org/10.1109/TBME.2015.2468589.

[11] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Eng Med Biol Mag. 2001;20:45–50.

[12] Castells F, Laguna P, Sörnmo L, Bollmann A, Roig JM. Principal component analysis in ECG signal processing. EURASIP J Adv Signal Process. 2007;2007:074580. https://doi.org/10.1155/2007/74580.

[13] Lyon A, Ariga R, Mincholé A, Mahmod M, Ormondroyd E, Laguna P, et al. Distinct ECG phenotypes identified in hypertrophic cardiomyopathy using machine learning associate with arrhythmic risk markers. Front Physiol. 2018;9:213. https://doi.org/10.3389/fphys.2018.00213.

[14] Mincholé A, Jager F, Laguna P. Discrimination between ischemic and artifactual ST segment events in Holter recordings. Biomed Signal Process Control. 2010;5:21–31. https://doi.org/10.1016/j.bspc.2009.09.001.

[15] Zihlmann M, Perekrestenko D, Tschannen M. Convolutional recurrent neural networks for electrocardiogram classification. Computing in Cardiology (CinC). 2017; 2017:1–4. https://doi.org/10.22489/CinC.2017.070-060.

[16] Martínez JP, Almeida R, Olmos S, Rocha AP, Laguna P. A wavelet-based ECG delineator: evaluation on standard databases. IEEE Trans Biomed Eng. 2004;51:570–81. https://doi.org/10.1109/TBME.2003.821031.

[17] Laguna P, Mark RG, Goldberg A, Moody GB. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. Comput Cardiol. 1997;1997:673–6. https://doi.org/10.1109/CIC.1997.648140.

[18] Camps J, Rodriguez B, Minchole A. Deep learning based QRS multilead delineator in electrocardiogram signals; 2018. https://doi.org/10.22489/CinC.2018.292.

[19] Zacur E, Minchole A, Villard B, Carapella V, Ariga R, Rodriguez B, et al. MRI-based heart and torso personalization for computer modeling and simulation of cardiac electrophysiology. Imaging for patient-customized simulations and Systems for Point-of-Care Ultrasound. Cham: Springer; 2017. p. 61–70. https://doi.org/10.1007/978-3-319-67552-7_8.

[20] Lyon A, Bueno-Orovio A, Zacur E, Ariga R, Grau V, Neubauer S, et al. Electrocardiogram phenotypes in hypertrophic cardiomyopathy caused by distinct mechanisms: apico-basal repolarization gradients vs. Purkinje-myocardial coupling abnormalities. Europace. 2018;20:iii102–12. https://doi.org/10.1093/europace/euy226.