Accepted Manuscript
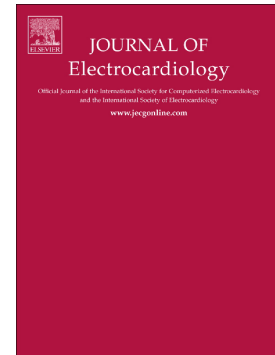
Readily available ECG databases

Nellyzeth Flores, Roberto L. Avitia, Marco A. Reyna, Conrado García

Please cite this article as: Nellyzeth Flores, Roberto L. Avitia, Marco A. Reyna, Conrado García , Readily available ECG databases. Yjelc (2018), doi:10.1016/j.jelectrocard.2018.09.012

Readily Available ECG Databases

Nellyzeth Flores[1], Roberto L. Avitia[1], Marco A. Reyna[2], Conrado García[2]

[1]Facultad de Ingeniería, Instituto de Ingeniería[2] Universidad Autónoma de Baja California, México

Email: {nellyzeth.flores, ravitia, cnrdgarciag, mreyna}@uabc.edu.mx

Abstract— Medical databases are an essential tool in health research that have become more frequently used in the last few decades alongside the growth and worldwide availability of the internet. Heart diseases are one of the most common health problems in the world with a vast amount of work being done on disease prevention and surveillance. With respect to arrhythmia detection, the electrocardiogram (ECG) is one of the most studied monitoring tools to date, generating large amounts of real-time data and giving rise to readily available ECG databases. In this research, we provide an overview of the most cited ECG databases with public/read access and discuss their characteristics with emphasis on the ECG records, as well as their use and applications carried out by other researchers around the world.

Keywords— cardiovascular disease, electrocardiogram, ECG databases.

INTRODUCTION

Nowadays, in the age of technology and network communication, the internet is an essential tool for carrying out almost any kind of work. The great advancements in medical research could not have been possible without simple and quick access to electronic resources. All modern technological tools have allowed for better and more fast-paced research in healthcare and biomedicine, including medical databases. Cardiovascular diseases (CVD) consist of pathologies that come from a multifactorial origin, and can occur at different ages in the human being, making treatment difficult in a variety of ways. The WHO and the Pan American Health Organization have highlighted CVD as the leading cause of premature deaths worldwide, marking them as red flags for attention due to a rising incidence of CVD in recent years [1].

METHODOLOGY

In order to undertake this mini literature review, we performed a literature search for electronic resources related to the subject of electrocardiography. We utilized several medical databases and ran a search query using very general and a few specialized keywords in order to capture a wide variety of articles on the subject. We used the *"IEEE xplore"*, *"SCOPUS"* and *"ScienceDirect"* search engines to find the electronic resources of interest for this review. We employed the keywords: "database", "electrocardiogram (ECG or EKG)", "cardiovascular disease (CVD)" and "acute myocardial infarction (AMI)". Furthermore, we utilized "AND" and "OR" connectors in order to create search chains that would better delimit the topic of interest. The inclusion criteria consisted of: articles and review papers published in journals from 01/01/2007 to 01/11/2017 and articles and review papers that utilized publicly available databases for cardiovascular diseases. The exclusion criteria consisted of: articles that were part of conferences, posters and book chapters as well as articles that studied medical images instead of ECG. A schematic diagram of the literature search is shown in Fig. 1.
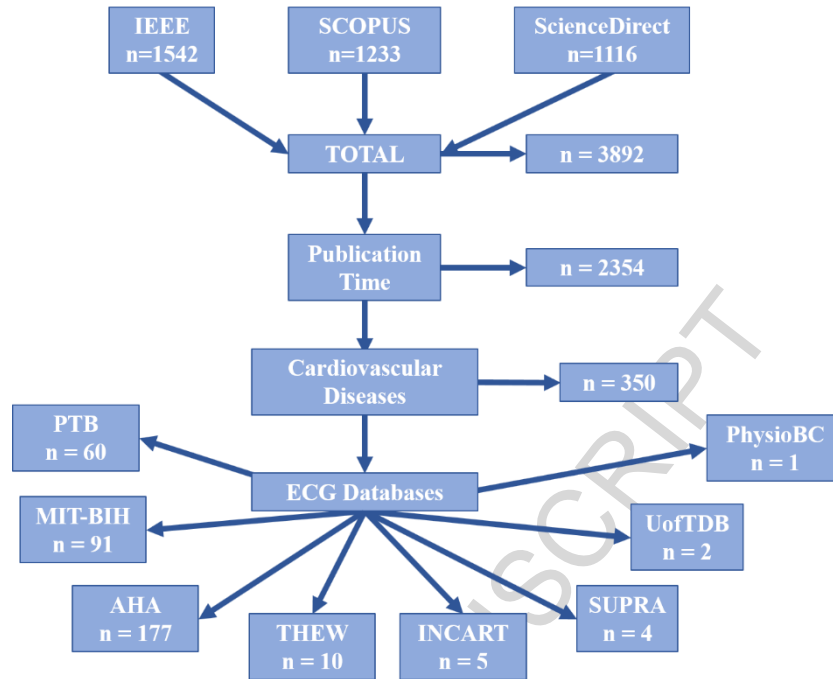
Fig. 1. ECG databases related with cardiovascular diseases

RESULTS

A total of 3892 articles were obtained from the 3 databases, with 2,354 articles found to be within the time period of interest. As a third filter, the phrase "cardiovascular diseases" was used, from which 350 articles were obtained. For more details about the ECG databases' description, see Table 1.

Table 1. Description of freely-available ECG databases

Database	Year	Number of subjects	Number of records	Length of the records	Resolution (bit)	Sampling rates	Number of ECG leads	Location of electrodes	Type of electrodes	Citations (2014 to 2017)
PTB	1995	290	549	N/A	16	10 kHz	12 + 3 Frank-lead	Chest, limbs	Gel, Ag/AgCl	60
MIT-BIH	1990	48	48	N/A	11	360 Hz	12	Chest, limbs	Gel, Ag/AgCl	91
AHA	1980	155	155	N/A	12	250 Hz	12	Chest, limbs	Gel, Ag/AgCl	177
THEW (AMI)	2012	93	160	N/A	12	200 Hz	3	Chest	Gel, Ag/AgCl	10
INCART	2007	32	75	30 minutes	12	257 Hz	12	Chest, limbs	Gel, Ag/AgCl	5
SUPRA	1991	78	78	30 minutes	12	125 Hz	12	Chest, limbs	Gel, Ag/AgCl	4
UofTDB	2014	1020	1020+	2 to 5 minutes	12	200 Hz	1 lead (DI)	Fingers	Dry Ag/AgCl	2
PhysioBC	2016	91	182	2 to 5 minutes	16 and 12	1 kHz	12 + 3 Frank-lead	Chest, limbs	Gel, Ag/AgCl	1

PTB Diagnostic ECG

The PTB (Physikalisch-Technische Bundesanstalt) database consists of 549 records obtained from 290 patients. The ages range from 17 to 87 years, providing detailed patient-level information, including: age, gender, diagnosis, positive medical history, medications, previous surgeries, existence of coronary artery disease or any other heart disease. This database uses the 12-leads ECGs and the Frank leads (x,y and z leads). The resolution is 16 bits, with 17 channels (15 for ECG, 1 for breathing and 1 for the voltage line), plus a noise voltage of 10 uV maximum. The PTB can be found on the website: https://www.physionet.org/physiobank/database/ptbdb/ [2] [3].

MIT-BIH Arrhythmia

This database is governed by the Harvard-MIT Division of Health Sciences and Technology within the Biomedical Engineering Center. It is composed of a collection of 48 half-hour records, each obtained from 47 patients between 1975 and 1979. These records were digitized using 360 samples per second per channel with an 11-bit resolution and a range of 10mV. The data contained in each patient file includes: date of registration, age, gender and prescribed medications. The database is hosted on: https://www.physionet.org/physiobank/database/mitdb/ [3] [4].

INCART – St. Petersburg Institute of Cardiological Technics

This database consists of 75 electrocardiography records, extracted from 32 Holter records, representing a total of 32 patients (17 men and 15 women) with ages ranging from 18 to 80 years and an average age of 58 years. None of the patients has an implanted pacemaker, but they have various cardiac problems, including myocardial ischemia, coronary artery disease, conduction abnormalities and arrhythmias. 12 leads ECGs were used with a sampling frequency of 257 samples per second and a duration of 30 minutes. The database is hosted on: https://www.physionet.org/pn3/incartdb/ [3].

SUPRA – Supraventricular Arrhythmia

This database was obtained in the MIT-Beth Israel Hospital (MIT-BIH) and contains 78 ECG records with a duration of half an hour, chosen as examples of the supraventricular arrhythmias. The records were obtained between 1990 and 1992, with subsequent annotations at 1999, 2010 and 2012. As methodology, HOBBES (Hypothetical Obliterated Beat Sequences) system was used [3]. It is available on: https://physionet.org/physiobank/database/svdb/.

AHA

This database can be obtained only from the emergency care Research Institute (ECRI Institute) website as a requested DVD, along with the payment of a fee. The DVD includes 154 records divided into 8 types of arrhythmias: Non-PVCs, Uniformly Isolated PVCs, Isolated Multiform PVCs, Bigeminus, R-T beats, Couplets, Ventricular rhythms and ventricular fibrillation or ventricular flutter. Each ECG record (one per patient) lasts 3 hours in total, divided in periods of at least 30 minutes, each for the beats classifications. The information was digitized in two ECG channels with 250 samples per second and 12 bits of precision. It is hosted in the link: www.ecri.org [5].

UofTDB – University of Toronto Database

This database consists of a sample of 1020 patients, whose single lead ECG was recorded from 2 - 5 minutes. Measurements were taken in different body postures (standing, sitting, tripod and supine) together with physical exercise. A subset of 43 patients attended 6 sessions for ECG recording over a period of 6 months. The leads were placed on the palms and fingers, obtaining the DI lead. It is hosted in the link: http://www.comm.utoronto.ca/~biometrics/databases.html [6].

THEW – Telemetric and Holter ECG Warehouse

THEW is an organization hosted by the Heart Research Follow-up Program at the Medical Center of the University of Rochester, with the project starting in 2008. It contains more than 3700 digitized ECG signals

mainly obtained through Holter-24 hours. These ECG records are part of 13 independent studies, which are available only to the members at www.thew-project.org. One of the studies, the Acute Myocardial Infarction Patients (AMI) consists of 90 patients, with 160 records, which were taken from 3 quasi-orthogonal leads. The sampling frequency is 200 samples per second with an amplitude of 10μV [7].

PHYSIOBC

This is a new database developed in Mexico. It consists of 91 patients and 182 records; the age ranges from 18 to 70 years. The sample rate for the 12- lead ECGs is 1000 samples per second, and for the Frank lead ECGs is 500 samples per second. The duration of the records is between 2 - 5 minutes per patient. The database is authorized by the Research Ethics Committee and evaluated by a cardiologist and is hosted on: www.physiobc.org [8].

CONCLUSION

This review has documented for the first time a variety of readily available ECG databases which facilitate further research in the area of electrocardiography.

Ethical approval: This article does not contain any studies performed by any of the authors on human participants or animals.

REFERENCES

[1]   Junell A, Thomas J, Hawkins L, Sklenar J, Feldman T, Henrikson CA, et al. Screening entire healthcare system ECG database : Association of deep terminal negativity of P wave in lead V1 and ECG referral with mortality. Int J Cardiol 2017;228:219–24. doi:10.1016/j.ijcard.2016.11.128.

[2]   Bousseljot R, Kreiseler D, Schnabel AN. The PTB diagnostic ECG database. Biomed Tech 1995;40:317. doi:10.13026/C28C71.

[3]   Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank,PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000:215–20. doi:10.1161/01.CIR.101.23.e215.

[4]   Moody GB MR. The impact of the MIT-BIH Arrhytmia Database. IEEE Eng Med Biol Mag 2001;20:45–50. doi:10.13026/C2F305.

[5]   Hermes, R. E., Geselowitz, D. B., & Oliver GC. Development, distribution, and use of the American Heart Association database for ventricular arrhytmia detector evaluation. Comput Cardiol 1980:263–6. doi:10.13026/C27P4P.

[6]   Pouryayevali Shahrzad, Wahabi Saeid, Hari Siddarth HD. On establishing evaluation standards for ECG biometrics. ICASSP 2014:3774–8. doi:10.1109/ICASSP.2014.6854307.

[7]   Couderc J. The Telemetric and Holter ECG Warehouse (THEW): The first three years of development and research ☆ J Electrocardiol 2012;45:677–83. doi:10.1016/j.jelectrocard.2012.08.001.

[8]   R. L. Avitia, Avena G, Flores N, Reyna MA, Nava ML. Physiological Data of Baja California: PhysioBC, Electrocardiographic Results 1, 2017;38:372–81. doi:dx.doi.org/10.17488/RMIB.38.1.33.