Norwegian Endurance Athlete ECG Database

Preprint · July 2022
DOI: 10.36227/techrxiv.20199731.v1

Bjørn-Jostein Singstad
Akershus universitetssykehus

Norwegian Endurance Athlete ECG Database

SUBMISSION DATE / POSTED DATE

30-06-2022 / 05-07-2022

DOI

10.36227/techrxiv.20199731.v1

IEEE Open Journal of
Engineering in Medicine and Biology

Norwegian Endurance Athlete ECG Database

Bjørn-Jostein Singstad*, *Member, IEEE*

Abstract*— Athletes often have training-induced remodeling of the heart, and this can sometimes be seen as abnormal but non-pathological changes in the electrocardiogram. However, these changes can be confused with severe cardiovascular diseases that, in some cases, can cause cardiovascular death. Electrocardiogram data from athletes is therefore important to learn more about the difference between normal athletic remodeling and pathological remodeling of the heart. This work provides a dataset of electrocardiograms from 28 Norwegian elite endurance athletes. The electrocardiograms are standard 12-lead resting ECGs, recorded for 10 seconds while the athlete's lay supine on a bench. The electrocardiograms were then interpreted by an interpretation algorithm and by a trained cardiologist. The electrocardiogram waveform data and the interpretations were stored in Python Waveform Database format and made publicly available through PhysioNet.

***Index Terms*— Electrocardiograms, athletes, dataset, physionet**

***Impact Statement*— This is the first open dataset with electrocardiograms recorded from athletes. With the recent advances in artificial intelligence-based ECG interpretation, this might be an important contribution towards future interpretation models.**

I. Introduction

ATLETES often have increased thickness in the left ventricular wall and extended chambers in both the left and right ventricle compared to untrained people at the same age. These structural changes can be difficult to distinguish between normal athletic remodeling of the heart and severe cardiovascular disease (CVD), such as hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), arrhythmogenic right ventricular cardiomyopathy (ARVC) and left ventricular noncompaction (LVNC). These pathologies are associated with a risk of Sudden Cardiac Death (SCD). Due to the low prevalence of these diseases, it requires very high accuracy in order to minimize the number of false positives.

Currently, there exist specific criteria for interpreting electrocardiograms (ECG) from athletes [1]. The interpretation criteria aim to detect athletes at risk of having a sudden cardiac arrest, but also point out fewer false positives. The latter is important because many false positives will represent a heavy burden on the health care system, and an unnecessarily high mental burden on the incorrectly interpreted individual athlete [2].

Submitted 30. June 2022
*Simula Research Laboratory, Department of Computational Physiology, Kristian Augusts gate 23, 0164, Oslo Norway. Email: bjornjs@simula.no.

The interpretation criteria for athletes are constantly evolving, and Berge et al. 2015 [3] showed how the Seattle criteria from 2013 [4] lowered the number of abnormal ECG findings from 29.3% (specified European Society of Cardiology recommendations [5]) to 11%. Furthermore, Refined Criteria (2014) [6] has been shown to lower the number of false positives further, and at the same time not lower the detection rate of sick athletes [1], [7]. On the other hand, it requires a lot of expertise to interpret these ECGs correctly, and in addition, these criteria are based on manual interpretation, which can be very time-consuming for the interpreter [8].

An alternative to manual interpretation is computer-based interpretation. A study compared the interpretation of ECGs from athletes using algorithms versus visual measurements performed by specialists and identified limitations of algorithm-based ECG interpretations on athletes [9]. On the other hand, new methods such as AI-based ECG interpretation have shown promising performance in the last couple of years [10]–[14] and these methods might improve today's ECG interpretation algorithms. In order to train AI-based algorithms, data is needed, and this paper presents the first open-access ECG database containing ECGs from elite endurance athletes. Therefore, this article marks the beginning of an open-source development of AI based ECG interpretation algorithms for athlete cohorts.

II. Data collection Procedures

A. Participants

The participants who donated their ECG to this study were informed and gave written consent before the data acquisition was initiated, they also agreed to have their ECG shared in an open database. The study protocol and consent form were approved by the Norwegian Centre for Research Data (application ID: 389013) and the University of Oslo. The ethical considerations were approved by the Regional Committees for Medical and Health Research Ethics (application ID: 51205).

Twenty-eight healthy athletes were recruited for this study. From Figure 1 we see that 19 (68%) of the participants were men and 9 (32%) were women. Participants' ages ranged from 20 to 43 years (Mean = 25 years, standard deviation = 4.7 years). The distribution among sports was 24 rowers (86%), 2 kayakers (7%) and 2 cyclists (7%). The average amount of training hours for 2017 was 822 hours with a standard deviation of ± 117 hours, in 2018 the average amount of training was 820 hours with a standard deviation of ± 113 hours and in 2019 the average amount of training was 798 hours with a standard deviation of ± 171 hours.

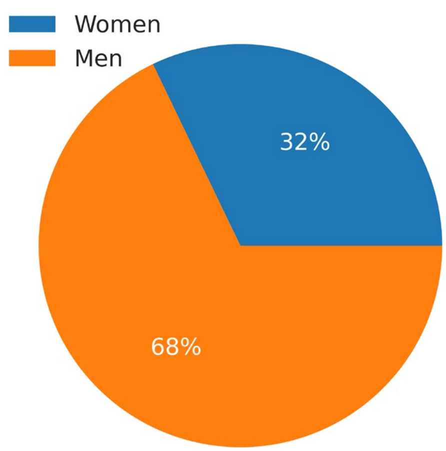

Fig. 1. The diagram shows the distribution between men and women in this dataset.

B. Signal Acquisition

The test subjects were lying horizontally on a bench, relaxing, while electrodes were attached to perform a 12-lead ECG recording. The precordial leads were attached to the chest as shown in Figure 3, and the limb leads were placed on the wrists and ankles as shown in Figure 2. The electrodes used were of the type Ambu© BlueSensorQ from Ballerup in Denmark.

The recordings were performed as a standard 10 seconds resting ECG and sampled with a sampling frequency of 500hz. The device used to record the ECGs was a GE MAC VUE 360.

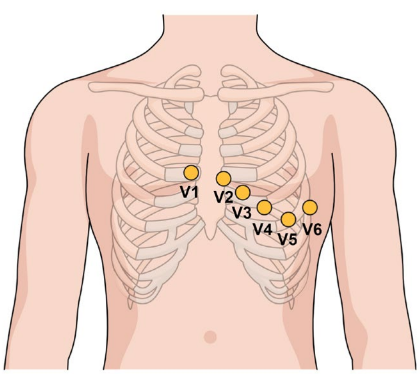

Fig. 3. The figure shows how the precordial leads were placed on the test subjects. The illustration is made using Mind the Graph (https://mindthegraph.com/).

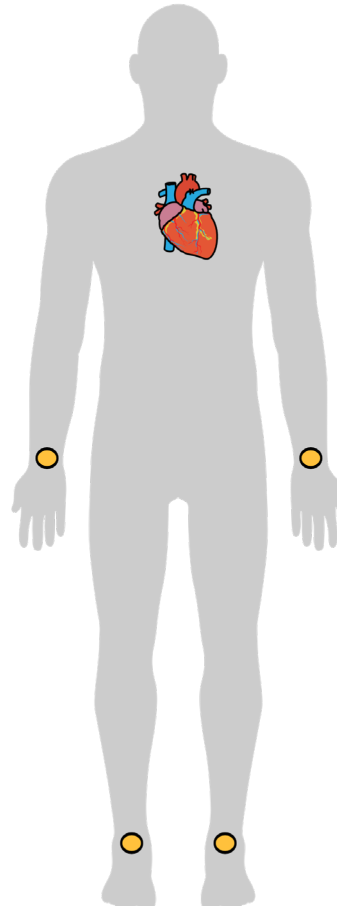

Fig. 2. The figure shows how the limb leads where placed on the the subjects in this study. The illustration is made using Mind the Graph (https://mindthegraph.com/).

C. Interpretation

GE MAC VUE 360's built-in interpretation algorithm, Marquette 12SL (version 23 (v243)), performed an automatic interpretation of all ECGs after each recording was taken. The interpretations given by the Marquette 12SL algorithm are summarized in Figure 4. Each ECG was automatically stored in the memory of the GE MAC VU360 device as a .ecg file format. The files were transferred via USB to a standalone PC where CardioSoft (Version V6.73) was used to convert the .ecg files to XML files. Furthermore, all ECG recordings were examined by a cardiologist, with a specialization in athletes' hearts. This cardiologist was given access to all the ECG recordings in PDF format. The PDF document include both signals from the 12 leads and interpretive texts from GE Marquette SL 12. The cardiologist interpreted the ECGs according to the international criteria for ECG interpretation of athletes [1] The interpretations given by the cardiologist are summarized in Figure 5.

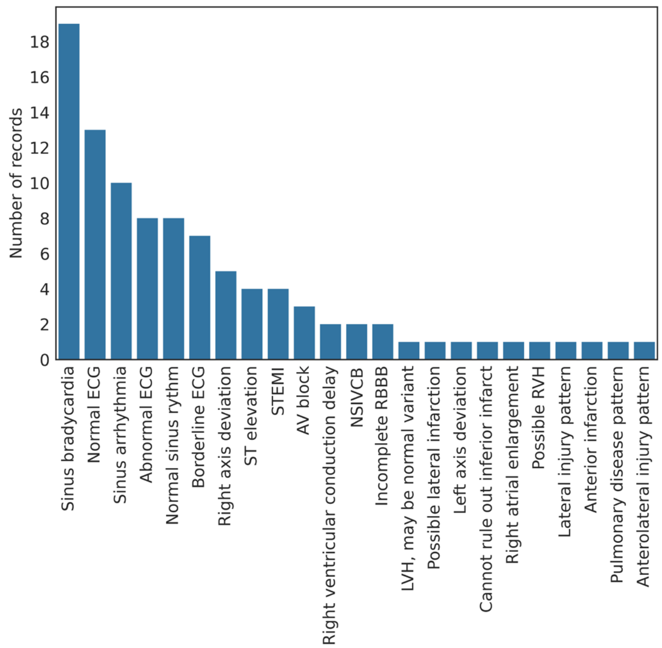

Fig. 4. The bar plot represents the prevalence of each diagnose (on the y-axis), given by the GE Marquette SL12 algorithm.

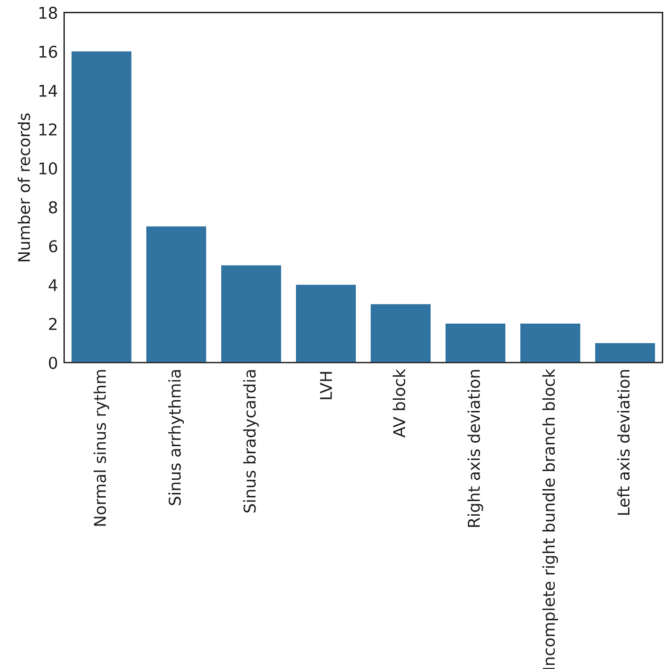

Fig. 5. The bar plot represents the prevalence of each diagnose (on the y-axis), given by the cardiologist.

III. Dataset Description

A. Data Records

The waveform data is given in a binary format without any modifications or additional filtering other than whats performed in the GE hardware. The ECG files were retrieved from the GE apparatus using a USB stick. The files were saved in an encrypted GE proprietary file format and in order to extract the raw ECG data GE CardioSoft was used to convert the .ecg files to XML files. The raw ECG waveform and the interpretation text from the built-in algorithm were then extracted from the XML files and stored in .dat and .hea files. The raw ECG recordings were stored in .dat -files with a corresponding .hea file containing all the metadata for the corresponding recording. These file formats are compatible with the WaveForm DataBase (WFDB) package and this makes it convenient to import the data to Python [15]. Each of the 28 .dat files consists of a 12 x 5000 array, where 12 is the number of leads and 5000 is the number of samples in each lead. The header file contains information about the total amount of leads, samples per lead and additional information about each lead. The last two lines in the header file contain the diagnosis given by the Marquette SL12 (SL12) algorithm and the cardiologist (C). An example of such a header file is shown in Table 1.

B. Technical Validation

The cardiologist who interpreted the ECGs also investigated the ECG waveforms to assess if there were misplaced electrodes. The cardiologist found misplaced electrodes in one case and this is pointed out in the interpretation text by the cardiologist in the .hea file.

IV. Usage Notes

The intended use of this database is for the development of better algorithms designed to make better diagnostics for athletes based on ECG. One of the unique features of this database is that the ECGs are annotated by both a trained cardiologist and by state-of-the-art ECG software.

TABLE I
HEADER FILE CONTAINING META DATA ABOUT THE CURRENT MEASURMNET AND THE TEST SUBJECT

ath_001 12 500 5000
ath_001.dat 16 50000/mV 16 0 10251 49595 0 I
ath_001.dat 16 50000/mV 16 0 -1096 35223 0 II
ath_001.dat 16 50000/mV 16 0 -10267 60826 0 III
ath_001.dat 16 50000/mV 16 0 -3724 3505 0 AVR
ath_001.dat 16 50000/mV 16 0 9391 26379 0 AVL
ath_001.dat 16 50000/mV 16 0 -5395 57481 0 AVF
ath_001.dat 16 50000/mV 16 0 13580 61759 0 V1
ath_001.dat 16 50000/mV 16 0 11410 33501 0 V2
ath_001.dat 16 50000/mV 16 0 14721 52508 0 V3
ath_001.dat 16 50000/mV 16 0 16103 51083 0 V4
ath_001.dat 16 50000/mV 16 0 6662 44197 0 V5
ath_001.dat 16 50000/mV 16 0 -3806 11333 0 V6
#SL12: sinus bradycardia with marked sinus arrhythmia, Right Axis Deviation, Borderline ECG
#C: Sinus arrhythmia, Normal ECG

An example of a metadata file. The first line of the table shows the recording number (A0001), number of ECG leads (12), sampling rate (500Hz), number of samples (7500), date (12-May-2020) and time (12:33:59). The next 12 rows again show the file name, each signal was written with 16 bit and 24 bit offset, the resolution of the voltage signal (1000 / mV), the ECG device's Analog-to-digital converter values (16 + 24), the basic value of the signal is 0 for all conductors, the first value of the signal, checksum and finally the name of the lead. In the last 6 rows, information is given about age, gender, diagnosis (Dx) coded / encrypted with SNOMED-CT code, Prescription (Rx), patient history (Hx) and symptom or operation (Sx).

Despite the measurements being taken from top-trained athletes, it is not confirmed whether they had athletic remodeling of the heart or not. No echocardiographic or other examinations were performed to investigate the structure of the heart.

The cardiologist who investigated the ECG concluded that nothing pathological was detected in the athlete's ECG. These athletes could therefore be considered to be a healthy cohort. Most of the open-source databases consist of ECGs from patients and elderly people. Training a machine learning model on such a population may cause the model to acquire bias from the dataset, and eventually make it more prone to fail outside the training population. In a case where one wants to classify healthy and sick based on ECG, adding the ECGs from the athletes to the healthy may give a more diverse data set and counteract bias.

V. Code Availability

The data set is available on the following webpage: https://physionet.org/content/norwegian-athlete-ecg/1.0.0/.

Acknowledgment

I will thank Professor Emeritus Knut Gjessdal for providing his medical expertise and interpreting all of the ECGs. This work was done at the University of Oslo and I will thank Professor Ørjan Grøttem Martinsen for providing appropriate facilities for ECG measurements.

Competing interests

The author declares no competing financial interests.

References

[1] J. A. Drezner *et al.*, ‘International criteria for electrocardiographic interpretation in athletes: Consensus statement’, *British Journal of Sports Medicine*, vol. 51, no. 9, pp. 704–731, May 2017, doi: 10.1136/bjsports-2016-097331.

[2] M. T. Stokstad, H. M. Berge, and K. Gjesdal, ‘Hjertescreening av unge idrettsutøvere’, *Tidsskrift for Den norske legeforening*, vol. 133, no. 16, pp. 1722–1725, 2013, doi: 10.4045/tidsskr.13.0016.

[3] H. M. Berge, K. Gjesdal, T. E. Andersen, E. E. Solberg, and K. Steine, ‘Prevalence of abnormal ECGs in male soccer players decreases with the Seattle criteria, but is still high’, *Scandinavian Journal of Medicine & Science in Sports*, vol. 25, no. 4, pp. 501–508, 2015, doi: 10.1111/sms.12274.

[4] J. A. Drezner *et al.*, ‘Electrocardiographic interpretation in athletes: the “Seattle Criteria”: Table 1’, *British Journal of Sports Medicine*, vol. 47, no. 3, pp. 122–124, Feb. 2013, doi: 10.1136/bjsports-2012-092067.

[5] Uberoi Abhimanyu *et al.*, ‘Interpretation of the Electrocardiogram of Young Athletes’, *Circulation*, vol. 124, no. 6, pp. 746–757, Aug. 2011, doi: 10.1161/CIRCULATIONAHA.110.013078.

[6] Sheikh Nabeel *et al.*, ‘Comparison of Electrocardiographic Criteria for the Detection of Cardiac Abnormalities in Elite Black and White Athletes’, *Circulation*, vol. 129, no. 16, pp. 1637–1649, Apr. 2014, doi: 10.1161/CIRCULATIONAHA.113.006179.

[7] J. A. Drezner, ‘18 highlights from the International Criteria for ECG interpretation in athletes’, *Br J Sports Med*, vol. 54, no. 4, pp. 197–199, Feb. 2020, doi: 10.1136/bjsports-2019-101537.

[8] M. Bickerton and A. Pooler, ‘Misplaced ECG Electrodes and the Need for Continuing Training’, *British Journal of Cardiac Nursing*, vol. 14, no. 3, pp. 123–132, Mar. 2019, doi: 10.12968/bjca.2019.14.3.123.

[9] H. M. Berge, K. Steine, T. E. Andersen, E. E. Solberg, and K. Gjesdal, ‘Visual or computer-based measurements: important for interpretation of athletes’ ECG’, *Br J Sports Med*, vol. 48, no. 9, pp. 761–767, May 2014, doi: 10.1136/bjsports-2014-093412.

[10] A. H. Ribeiro *et al.*, ‘Automatic diagnosis of the 12-lead ECG using a deep neural network’, *Nature Communications*, vol. 11, no. 1, Art. no. 1, Apr. 2020, doi: 10.1038/s41467-020-15432-4.

[11] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, ‘Multi-class Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network’, *Information Fusion*, vol. 53, pp. 174–182, Jan. 2020, doi: 10.1016/j.inffus.2019.06.024.

[12] D. Li, H. Wu, J. Zhao, Y. Tao, and J. Fu, ‘Automatic Classification System of Arrhythmias Using 12-Lead ECGs with a Deep Neural Network Based on an Attention Mechanism’, *Symmetry*, vol. 12, no. 11, Art. no. 11, Nov. 2020, doi: 10.3390/sym12111827.

[13] T.-M. Chen, C.-H. Huang, E. S. C. Shih, Y.-F. Hu, and M.-J. Hwang, ‘Detection and Classification of Cardiac Arrhythmias by a Challenge-Best Deep Learning Neural Network Model’, *iScience*, vol. 23, no. 3, p. 100886, Mar. 2020, doi: 10.1016/j.isci.2020.100886.

[14] Z. I. Attia, D. M. Harmon, E. R. Behr, and P. A. Friedman, ‘Application of artificial intelligence to the electrocardiogram’, *European Heart Journal*, vol. 42, no. 46, pp. 4717–4730, Dec. 2021, doi: 10.1093/eurheartj/ehab649.

[15] Moody, George, Pollard, Tom, and Moody, Benjamin, *WFDB Software Package*. PhysioNet. doi: 10.13026/ZZPX-H016.