ANALYSIS AND DETECTION OF LATENT RHEUMATIC HEART DISEASE FROM LOW-COST ELECTROCARDIOGRAM SIGNALS USING DEEP LEARNING

DMP-RHDML ZB/22/032

ADMIN DETAILS

Project Name: Zuidbeurzen 2022 Referentie nr.: ZB/22/032 - Chuma Amsalu Tomas

Project Identifier: ZB/22/032 (BOF)

Grant Title: Analysis and Detection of Latent Rheumatic Heart Disease from Low-cost Electrocardiogram Signals using Deep Learning

Principal Investigator / Researcher: Prof. Dr. Rik Willems, Prof. Ir. Bart Vanrumste

Project Data Contact: Amsalu Tomas Chuma, +32 486 76 0 953, amsaluthomas.chuma@kuleuven.be

Description: Rheumatic heart disease (RHD) is an endemic disease in the global south and highly prevalent among school children, contributing a huge burden to national health. The disease is caused by strep-throat from beta-hemolytic group-A streptococcal infection, and progressively leading to severe cardiac valve damages when left untreated. If the disease is detected at early stage, it only requires prophylaxis of benzathine penicillin G (BPG) which can easily be administered at nearby local health centers. The gold standard diagnosis criteria by the World Heart Federation (WHF) dictates the use of echocardiography screening. However, poor and low resource health center settings and limited number of cardiologists in the global south worsen the problem. As a consequence, the WHF criteria has been impractical for a holistic intervention of the disease at the target communities. Conversely, affordable low-cost diagnosis studies for RHD are non-existent despite epidemiological studies in different low-income countries. Therefore, this study aims to exploit the use of low-cost phonocardiogram (PCG) and electrocardiogram (ECG) sensors for detection of latent RHD using an artificial intelligence (AI) algorithm. The dataset will be collected from patients living at prone areas and age-groups in Ethiopia. The outcomes of the research could result in an automated affordable RHD screening mechanism, and contributes to possible recommendations on RHD diagnostic criterions, RHD vaccine development, and medical marketing ventures in low-income countries. For this study, 450 participants (10–20 years of age) will be recruited from three different schools. The data to be collected and examined consists of cardiac auscultation using digital stethoscope, electrocardiogram using Kardiamobile 6L and echocardiograph images recorded using handheld ultrasound.

Institution: KU Leuven

1. GENERAL INFORMATION

a. Name applicant

Amsalu Tomas Chuma

b. BOF Project Number & Title

BOF project reference number: ZB/22/032

Title: Analysis and Detection of Latent Rheumatic Heart Disease from Low-cost Electrocardiogram Signals using Deep Learning

c. Affiliation

- KU Leuven

2. DATA DESCRIPTION

a. Will you generate/collect new data and/or make use of existing data?

- Collect new data

b. Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project.

Type of Data	Format	Volume	How created
(A) Electrocardiogram (ECG)	.mat, .hdf5	5 GB	The electrical activity of the heart will be recorded using KardiaMobile ECG sensor. The ECG signals from the electrodes on the KardiaMobile sensor transmit via Bluetooth connectivity to a paired device with a Kardia app installed. The Kardia app of the sensor works both in iOS and android phones to record the ECG signals. The ECG data will be recorded for 1 minute from each participant.
(B) Cardiac auscultation or phonocardiogram (PCG)	.m(at), .aac, .wav, .csv	10 GB	From each participant a 1 minute phonocardiogram data of the heart auscultation will be recorded using ThinkLabsOne digital stethoscope. This stethoscope can be connected to a smartphone via a cable to listen and record the auscultations in real-time. There is no additional software required in order to save the recordings. Hence, the standard voice record functionality in all iOS or Android smartphones suffices.
(C) Cardiac echo images	.JPEG, DICOM, .SPAR, .SDAT	35 GB	Cardiac ultrasound images will be collected from both RHD positive and RHD negative participants. The images will be taken mainly from PLAX view both with color flow and without color flow. Additional images can also be taken from apical 3-chamber view and PSAX views.
(D) Demographic data	.txt, .xlsx, .docx	0.5 GB	• .txt documents containing information about diagnosis (ground truth) label, medical history • .xlsx records of list of participants and their demographic data. All paper records will be

			transferred to digital .xlsx format as soon as possible • .docx for instructions on data collection procedures, adverse events, screening task evaluation reports
(E) Analysis scripts and code for echo images, and statistical analysis	.m(at), .py(w), , .jmp, .dll	2 GB	Programming codes to preprocess the dataset and its statistics
(F) Processed data	.xlsx, .txt, .csv	0.5 GB	Processed data of all raw data sources described above will be stored in .xlsx, .txt format for diagnosis labels, medical history and demographics. The source medical data to be analyzed will be stored in .JPEG, .wav/.mat/.hdf5 formats depending on the type of data
(G) Metadata	.txt/.docx	100 MB	see below, section 4 (b)

Table 1: Data types and formats

3. LEGAL AND ETHICAL ISSUES

a. Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- Yes

Privacy Registry Reference: G-2022-5838 (accepted)

Short description of the kind of personal data that will be used: Source data to be collected are participant's demographics, body mass index (BMI), medical history, ECG data for a brief period of 30 seconds twice, PCG or auscultation data for a brief period of 30 seconds twice, and echocardiograph images. The source data will be collected and recorded in the study participant's files/medical records. Furthermore, all personal information will be pseudonymized before starting the medical examination.

b. Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

KU Leuven PRET reference: G-2022-5838 (accepted)
Ethical commission reference: S67064 (pending)

c. Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No. But if a need arises for a possible IP restrictions, then proper communication will be made to LRD.

d. Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. DOCUMENTATION AND METADATA

a. What documentation will be provided to enable reuse of the data collected/generated in this project?

First of all, a training on data collection procedures, data standards and how/where to store personal or sensitive data, and mechanisms to handle the data according to the GCP-R2 guidelines will be provided for new authorized research members who want to reuse the data. Guidelines and instructions can be availed as .pdf and/or .readme; and it will be stored on network drive. In addition, all the information regarding this study should be kept on central secured Drives of KU Leuven. It needs to be updated by a member of the research team when a new data collection from participants at the study site take place. Further details such as study goal, objectives of the study and practical matters to replicate the trial are found in the study protocol. There will also be a git repository to post all the codes for the project, procedures to compile/run to reproduce the reported results.

b. Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

● Yes

Letters in brackets below refer to the Table-1 above (question 2b).

(A) and (B): Raw ECG and PCG data (.mat, .csv, .hdf5) will be stored for each condition with a .txt README file explaining the data labels and data acquisition parameters and lead systems. Furthermore, WFDB (Waveform Database) file formats will be maintained thereby additional files such as WFDB annotation file formats (annot), WFDB header file format(header), WFDB signal file formats (the signal), and WFDB calibration file format (wfdbcal) will be kept in the data folder.

(C): The cardiac ultrasound images will be used according to Digital Images and Communications in Medicine (DICOM) format (https://dicom.nema.org/medical/dicom/current/output/chtml/part10/chapter_7.html). An additional .docx file explaining the RHD screening protocol, echo image saving, data storage, primary data processing and generic descriptions of the applied data analysis processes will also be stored.

(D): For demographic data, personal details of each participant will be stored in .xlsx/.csv file. This file will be exported to RedCap to be stored in a structured manner in the data storage derives of KU Leven. Medical history and diagnosis description including the ground truth labels will be stored as a .txt file. This file will also contain information about diagnosis, device used, time where the diagnosis is performed, other key findings, adverse events if occur, and parent storage directory path information. The signed consent forms and other paper works will be stored in .pdf file format.

(E) and (F): Scripts will use the comments to explain each analysis step. The dataset and .xlsx documents will have a clear document name and row/column description; if needed for better clarification, further metadata (.txt/.docx) will be created.

Overall, for organizing and storing medical data a standard structures will be used. For physiological data (ECG/PCG) WFDB and .csv will be used in general. Whereas DICOM will be used for storing cardiac echo images. Every single record will have a metadata and header information containing data acquisition details (sampling rate, duration, axis, diagnosis label, etc.) will be included. For demographic and medical history data can be used as described on (D) above.

5. DATA STORAGE AND BACKUP DURING THE FWO PROJECT

a. Where will the data be stored?

   I. The time-stamped master copy of the data will be kept on the different local storage facility (OneDrive). Copies can be made and kept on password protected work computers/drives if needed for analyses/transfer.

  II. Since we will be working with sensitive personal data, data will be pseudonymized as soon as possible. Only one record that is linking the pseudonym to the personal data ('Participant identification list') will be kept on a second separate drive, that will be password secured. Access will be granted to researchers directly involved in the maintenance of this database and be kept as limited as possible.

  III. Since a full anonymization of data is not possible, all data will be kept on a password secured and encrypted online drive described above (see (I)), audited and declared suitable for storing medical data. Access has to be granted to each involved researcher separately by the principal investigator.

  IV. Physical data such as signed consent forms will be stored in a locked filing cabinet in an office at Groep-T, KU Leuven, 3000 Leuven.

b. How is backup of the data provided?

The data will be stored on different local storage with automatic daily back-up procedures that allow for disaster recovery. This can be implemented according to the general ICT back-up Policy.

c. Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available then explain how this will be taken care of.

- Yes

KU Leuven Enterprise Drive allows for unlimited data storage and separate Drives (e.g., pseudonymized data apart from personal data) for other users if needed. The backup capacity is technically not an issue.

d. What are the expected costs for data storage and back up during the project? How will these costs be covered?

Costs of data storage within KU Leuven using the default storage servers are borne centrally by KU Leuven hence no costs.

e. Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Please see question 5a. Furthermore, data will be encrypted on the level of the drive, and identifiable and pseudonymized data will be stored on 2 separate drives.

6. DATA PRESERVATION AFTER THE FWO PROJECT

a. Which data will be retained for the expected 5-year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

After the end of the project, all data will be retained for the 5-year period expected by KU Leuven. Identifying personal data will then be removed, for other data please see question 6b.

b. Where will the data be archived (= stored for the longer term)?

The data will be stored on the KU Leuven servers (with automatic back-up procedures) for 25 years, that conform the RDM policy.

c. What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

None. As explained in section 5, costs for data storage will be borne centrally by KU Leuven.

7. DATA SHARING AND REUSE

a. Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

b. Which data will be made available after the end of the project?

Data will only be made available in case of publications that require the publication/disclosure of the dataset. Because of the nature of medical images data that does not allow for full anonymization, even when removing all personal information from the files and defacing the images, this will be kept restricted.

In case data sharing is planned in the context of a publication, the privacy experts of KU Leuven will be consulted prior to publication to conform with all current privacy standards.

c. Where/how will the data be made available for reuse?

- Upon request by mail

As explained above, medical data is sensitive data. Therefore, re-use within or outside of the research group will be provided if requested via mail. In this case, only the necessary pseudonymized information will be shared. In case of data sharing outside of the eMedia research groups of KU Leuven joint data sharing agreement should be filed and the universities' privacy and legal experts will be consulted prior to data sharing to conform with all current privacy standards and regulate the data sharing process.

d. When will the data be made available?

- Upon publication of the research results

Data will only be made available to other researchers after publication of the research results.

e. Who will be able to access the data and under what conditions?

As stated above, only requests via mail will be answered. Privacy and legal experts will be consulted when sharing data with researchers outside of the research group. A written agreement with the PI is necessary when sharing the data outside of the eMedia research group.

f. What are the expected costs for data sharing? How will the costs be covered?

None. Data preparation (defacing, removal of personalized data in the imaging files, etc.) will be done by the researchers primarily involved in the project. Secure data sharing infrastructure is available at KU Leuven for free, e.g. Belnet. If costs incurred, these need to be covered by the requesting party(-ies).

8. RESPONSIBILITIES

a. Who will be responsible for data documentation & metadata?
The researcher, when his contract has ended the responsibility shifts to Prof. Ir. Bart Vanrumste to ensure data preservation and reuse.

b. Who will be responsible for data storage & back up during the project?
The researcher, when his contract has ended the responsibility shifts to Prof. Ir. Bart Vanrumste to ensure data preservation and reuse.

c. Who will be responsible for ensuring data preservation and reuse?
The researcher, when his contract has ended the responsibility shifts to Prof. Ir. Bart Vanrumste to ensure data preservation and reuse.

d. Who bears the end responsibility for updating & implementing this DMP?
The PI bears the end responsibility of updating & implementing this DMP.