South Africa

Professor Deenan, Pillay, UCL; AHRI
Dr Olivier, Koole, LSHTM; AHRI
Dr. Emily, Wong, AHRI

Performance of computer-aided chest radiography for detection of TB in a community-based multimorbidity cohort in rural KwaZulu-Natal

Study Documentation

June 8, 2022

Metadata Production

Metadata Producer(s)	Africa Health Research Institute (AHRI)
Identification	DDI.AHRI.Vukuzazi.ChestXrayPaperData

Table of Contents

Performance of computer-aided chest radiography for detection of TB in a community-based multimorbidity cohort in rural KwaZulu-Natal

Overview

Identification	AHRI.Vukuzazi.ChestXrayPaperData
Version	V1.0.0

Abstract
The datasets provided here will be used for devising a machine learning tool to improve upon CAD4TB's triaging of chest x-rays.

Kind of Data	Survey Data
Unit of Analysis	Each chest x-ray image or images are for a Vukuzazi study participant who had chest x-ray done. The clinical data, each record is for a single participant who participated in Vukuzazi and had chest x-ray done. Some individuals had more than one chest x-ray images and therefore have more than one record.

Scope & Coverage

Keywords	Tuberculosis
Topics	Chest x-ray; machine learning; tuberculosis; population health screening
Time Period(s)	2018-2019
Countries	South Africa

Geographic Coverage
uMkhanyakude district in northern KwaZulu-Natal

Universe
Individuals aged 15 years and above who are resident members of households in the AHRI DSS area.

Producers & Sponsors

Primary Investigator(s)	Professor Deenan, Pillay, UCL; AHRI Dr Olivier, Koole, LSHTM; AHRI Dr. Emily, Wong, AHRI
Other Producer(s)	Africa Health Research Institute (AHRI)
Funding Agency/ies	Wellcome Trust (WT) , Core funding
Other Acknowledgment(s)	Dr. Kobus, Herbst , Chief Information Officer , AHRI Dickman, Gareta , Head of Research Data Management , AHRI Jaco, Dreyer , Data management, data cleaning , AHRI Tshwaragang, Modise , Data management, data cleaning , AHRI Siyabonga, Nxumalo , Data management, data cleaning , AHRI Sweetness, Dube , Data documentalist , AHRI

Sampling

Sampling Procedure

N/A. All individuals aged 15 years and older who are resident members of households in the DSS area are eligible to participate Vukuzazi study.

Data Collection

Data Collection Dates	start 2018-05-01 end 2019-05-31

Data Processing & Appraisal

Data Editing
NA

Accessibility

Access Conditions

The representative of the Receiving Organization agrees to comply with the following conditions:

1. Access to the restricted data will be limited to the Lead Researcher and other members of the research team listed in this request.
2. Copies of the restricted data or any data created on the basis of the original data will not be copied or made available to anyone other than those mentioned in this Data Access Agreement, unless formally authorized by the Data Archive.
3. The data will only be processed for the stated statistical and research purpose. They will be used for solely for reporting of aggregated information, and not for investigation of specific individuals or organizations. Data will not in any way be used for any administrative, proprietary or law enforcement purposes.
4. The Lead Researcher must state if it is their intention to match the restricted microdata with any other micro-dataset. If any matching is to take place, details must be provided of the datasets to be matched and of the reasons for the matching. Any datasets created as a result of matching will be considered to be restricted and must comply with the terms of this Data Access Agreement.
5. The Lead Researcher undertakes that no attempt will be made to identify any individual person, family, business, enterprise or organization. If such a unique disclosure is made inadvertently, no use will be made of the identity of any person or establishment discovered and full details will be reported to the Data Archive. The identification will not be revealed to any other person not included in the Data Access Agreement.
6. The Lead Researcher will implement security measures to prevent unauthorized access to licensed microdata acquired from the Data Archive. The microdata must be destroyed upon the completion of this research, unless the Data Archive obtains satisfactory guarantee that the data can be secured and provides written authorization to the Receiving Organization to retain them. Destruction of the microdata will be confirmed in writing by the Lead Researcher to the Data Archive.
7. Any books, articles, conference papers, theses, dissertations, reports, or other publications that employ data obtained from the Data Archive will cite the source of data in accordance with the citation requirement provided with the dataset.
8. An electronic copy of all reports and publications based on the requested data will be sent to the Data Archive.
9. The original collector of the data, the Data Archive, and the relevant funding agencies bear no responsibility for use of the data or for interpretations or inferences based upon such uses.
10. This agreement will come into force on the date that approval is given for access to the restricted dataset and remain in force until the completion date of the project or an earlier date if the project is completed ahead of time.
11. If there are any changes to the project specification, security arrangements, personnel or organization detailed in this application form, it is the responsibility of the Lead Researcher to seek the agreement of the Data Archive to these changes. Where there is a change to the employer organization of the Lead Researcher this will involve a new application being made and termination of the original project.
12. Breaches of the agreement will be taken seriously and the Data Archive will take action against those responsible for the lapse if willful or accidental. Failure to comply with the directions of the Data Archive will be deemed to be a major breach of the agreement and may involve recourse to legal proceedings. The Data Archive will maintain and share with partner data archives a register of those individuals and organizations which are responsible for breaching the terms of the Data Access Agreement and will impose sanctions on release of future data to these parties.

Citation Requirements

Performance of computer-aided chest radiography for detection of TB in a community-based multimorbidity cohort in rural KwaZulu-Natal - Overview

- 6 -

Pillay, D., Koole, O. and Wong, E. (2020) "Performance of computer-aided chest radiography for detection of TB in a community-based multimorbidity cohort in rural KwaZulu-Natal." Africa Health Research Institute (AHRI). doi: 10.23664/AHRI.VUKUZAZI.CHESTXRAYPAPERDATA

Performance of computer-aided chest radiography for detection of TB in a community-based multimorbidity cohort in rural KwaZulu-Natal - Overview

- 6 -

Files Description

Dataset contains 1 file(s)

public_release	
# Cases	9912
# Variable(s)	12

Performance of computer-aided chest radiography for detection of TB in a community-based multimorbidity cohort in rural KwaZulu-Natal - Variables List

- 8 -

Variables List

Dataset contains 12 variable(s)

File public_release

#	Name	Label	Type	Format	Valid	Invalid	Question
1	Individu ..	IndividualId	continuous	numeric.0	9912	0	-
2	Sex	Sex	discrete	character-6	9912	0	-
3	age_group	age_group	discrete	character-5	9912	0	-
4	any_symp	any_symp	discrete	numeric.0	9912	0	-
5	binary_HIV	binary_HIV	discrete	character-12	9912	0	-
6	GeneXpert	GeneXpert	discrete	character-2	9912	0	-
7	LiquidCu ..	LiquidCulture	discrete	character-54	9912	0	-
8	Radiolog ..	RadiologistLungField	discrete	character-20	9912	0	-
9	CxrRadio ..	CxrRadiologistConclusion	discrete	character-62	9912	0	-
10	v5_CXRSc ..	v5_CXRScore	continuous	numeric.0	9912	0	-
11	v6_CXRSc ..	v6_CXRScore	continuous	numeric.0	9912	0	-
12	v7_CXRSc ..	v7_CXRScore	continuous	numeric.2	9912	0	-

Performance of computer-aided chest radiography for detection of TB in a community-based multimorbidity cohort in rural KwaZulu-Natal - Variables Description

- 9 -

Variables Description

Dataset contains 12 variable(s)

File : public_release

IndividualId: IndividualId

Information	[Type= continuous] [Format=numeric] [Range= 5-21428] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-] [Mean=10272.074 /-] [StdDev=6200.214 /-]

Sex: Sex

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-]

age_group: age_group

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-]

any_symp: any_symp

Information	[Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-]

binary_HIV: binary_HIV

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-]

GeneXpert: GeneXpert

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-]

LiquidCulture: LiquidCulture

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-]

RadiologistLungField: RadiologistLungField

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-]

CxrRadiologistConclusion: CxrRadiologistConclusion

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-]

v5_CXRScore: v5_CXRScore

Information	[Type= continuous] [Format=numeric] [Range= 10-100] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-] [Mean=33.742 /-] [StdDev=16.102 /-]

v6_CXRScore: v6_CXRScore

Information	[Type= continuous] [Format=numeric] [Range= 0-99] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-] [Mean=33.403 /-] [StdDev=18.9 /-]

v7_CXRScore: v7_CXRScore

Information	[Type= continuous] [Format=numeric] [Range= -2-98.95] [Missing=*]
Statistics [NW/ W]	[Valid=9912 /-] [Invalid=0 /-] [Mean=17.985 /-] [StdDev=17.782 /-]