J Nepal Health Res Counc 2024 Jul.-Sept.;22 (64): 477-483

Diagnostic Accuracy of Artificial Intelligence-Based Chest X-Ray reading for screening of Tuberculosis

Ruby Maka Shrestha,[1] Ram Chandra Paudel,[1] Puspanjali Adhikari,[1] Ram Hari Ghimire,[2] Karma Gurung,[3] Rajeev Shrestha[1]

[1]Dhulikhel Hospital, Kathmandu University Hospital, Dhulikhel, Kavre, Nepal, [2]Nobel Medical College and Teaching Hospital, Biratnagar, Morang, Nepal, [3]International Organisation for Migration, Baluwatar, Kathmandu, Nepal.

ABSTRACT

Background: Tuberculosis remains a public health challenge in Nepal and ranks as the seventh leading cause of death in the country. The END Tuberculosis strategy stresses - the screening for symptoms alone may not suffice; additional screening tools such as a chest radiograph may facilitate referral for diagnosis of tuberculosis. The study aims to evaluate the diagnostic accuracy of artificial intelligence (AI) based Chest X-ray and compare it with the human reading (radiologist), using GeneXpert-MTB RIF Assay for tuberculosis case detection.

Methods: Tuberculosis-suspected patients with a history of cough were screened using chest X-rays at two study sites (Dhulikhel Hospital and Nobel Medical College). The reading of AI qXR software was compared with radiologists reading who were blinded of the results generated by the software.

Results: The sensitivity of the test by qXR-based AI reading was 100%, (95% CI: 40 – 100%) and specificity 80% (95% CI: 73 – 87%), whereas the sensitivity of the test by the radiologist was 100%, (95% CI: 40 – 100%); and specificity 62% (95% CI: 53 – 70%).

Conclusions: Higher sensitivity and specificity were observed for both qXR-based AI and Radiographer readings for the diagnosis of TB.

Keywords: Artificial Intelligence; chest X-ray; diagnostic accuracy; screening; tuberculosis.

INTRODUCTION

Tuberculosis (TB) remains a public health challenge in Nepal and ranks as the seventh leading cause of death in the country.[1,2] In 2022, an estimated 10.6 million people worldwide were infected with tuberculosis, and a total of 1.3 million deaths were attributed to TB, making it the second leading mortality from infectious disease after COVID-19.[3] The World Health Organization (WHO) developed the 'End TB Strategy' which calls for countries to end the global TB burden to the levels envisioned by 2035.[4]To achieve this goal, it is essential to focus on active case-finding[5] and the adoption of innovative approaches, such as digital health to enhance patient care, surveillance, program management, training and communication.[6] In the high burden contexts like Nepal, the dependency on the passive case-finding approach should be shifted and more proactive TB screening should be strengthened.[7,8] Hence, the study was planned to pilot the effectiveness of Artificial Intelligence (AI)-enabled Chest X-ray (CXR) as a screening tool in TB case diagnosis at tertiary hospitals in Nepal. The study aimed to evaluate the outcome in terms of increased screening accuracy for enhancing quicker decision-making during TB treatment and determined the feasibility of replicating it in rural health facilities of Nepal.

METHODS

A cross-sectional pilot study was designed to evaluate the diagnostic accuracy of AI-based CXR. The result of the AI-based reading of the qXR was compared with the

Correspondence: **Prof Dr Rajeev Shrestha, Dhulikhel Hospital, Kathmandu University Hospital, Dhulikhel Kavre, Nepal, Email: rmaleku@hotmail.com.**

radiologist reading. However, the GeneXpert-MTB RIF Assay test was considered a standard TB case detection tool in this study.

qXR is a CXR screening tool that detects signs of pulmonary, hilar, and pleural tuberculosis. The artificial intelligence algorithm underlying qXR is trained to detect classical primary pulmonary TB. The AI algorithm is deployed within a complete workflow management platform that allows users to register and track patients through the process of clinical and X-ray screening.[9]

GeneXpert MTB-Rif is innovative semi-automated real-time polymerase chain reaction (PCR) nucleic acid amplification technology, which can simultaneously detect Mycobacterium Tuberculosis (MTB) and Rifampicin resistance (RIF).[10] It has shown high sensitivity (88%), specificity(99%) and a negative predictive value greater than 98% for TB detection.[11]

Primary data was collected at two study sites: Dhulikhel Hospital Kathmandu University Hospital (DHKUH) and Nobel Medical College and Teaching Hospital (NMCTH), both affiliated with Kathmandu University School of Medical Sciences, located in Bagmati Province and Koshi Province respectively. The data was collected from 20 December 2021 to 13 March 2022 at DHKUH, whereas at NMCTH, data were collected between 19 January 2022 and 9 March 2022 respectively.

Patients who visited the outpatient clinics of both the internal medicine departments and pulmonary units at both institutions were screened for TB. Screening was done for symptoms of TB which included a history of cough for a week or more, fever, weight loss, night sweats, and hemoptysis. In this study, we included adult patients aged more than 18 years. We also included patients with comorbidities such as diabetes, cancer, HIV infection, exposure to drug-resistant TB and Pulmonary Bacteriologically Confirmed (PBC) TB among family members, and those on long-term steroid therapy for further TB screening. The patients with extra-pulmonary tuberculosis, current users of anti-TB medication, and seasonal migrants with less than six months of residency were excluded from the study.

In the formative phase, we visited both hospitals to inquire about the readiness and willingness of the hospitals to participate in the study. A clinical team including two Radiologists with prior experience in interpreting chest X-rays, a Radiographer with previous experience in conducting a chest X-ray and a Lab technician with experience in conducting GeneXpert tests in the respective hospitals were purposively assigned by both hospitals for the study. The non-medical team included a research project coordinator, research assistants with backgrounds in public health and an IT technician experienced in medical software applications within a hospital were also assigned for both hospitals. Following the team assignment, an IT specialist from IOM and a representative from the Qure.ai team conducted separate seminars at each hospital. Qure.ai is an AI-based team that developed qXR software used in this study for reporting chest X-rays. The same technicians led both seminars. Following the seminar, both hospitals were equipped with an AI setup (qXR software) which enabled the automated chest reading (computerized chest radiography, CXR) of the patients. A week-long pretest was conducted at each hospital to ensure the research study's workflow was as planned and to confirm the system's functionality.

The research assistants collected basic demographic information about the patients and developed a personal code for each patient to maintain the confidentiality of the patients. Each patient received posterior-anterior CXR using digital X-ray machines. Each CXR was classified as "TB presumptive" if any pulmonary abnormality was detected by human readers, regardless of the abnormality being TB-specific, active or old. The results generated by machine learning for TB cases blinded every radiograph read by the radiologists. However, the final confirmation of the test result was considered based on GeneXpert's diagnosis of the sputum sample. GeneXpert diagnosis was therefore used as the reference for the study. The results of clinical investigation by the Radiologist were collected by the Research Assistant and hence updated to the online database.

The data was imported to R software (version 4.1.3) for the data analysis. The outcome of this study was an evaluation of the performance of AI in the interpretation of the CXR and a comparison of the result of the index test with the GeneXpert diagnosis to establish the presence or absence of TB in the study. Similarly, the diagnostic accuracy of AI and human reading in detecting TB was weighed against each other in terms of specificity and sensitivity and the positivity yield guided the recommendation made to the National Tuberculosis Control Center (NTCC) on the effectiveness of AI over human reading.

The study protocols were reviewed and approved by the Institutional Review Board at the KUSMS (IRC-KUSMS Approval No.: 238/2021), and the Nepal Health Research Council (NHRC) (Ref. No.:685). Data was stored in the cloud and kept confidential, which was made available only to the relevant research team.

RESULTS

A total of 168 patients participated in the study, with equal representation of both Male (50%) and female (50%). During an initial screening, the majority of patients did not report the presence of any co-morbidities including smoking, diabetes, hypertension, cancer, respiratory condition and old age. Only 11% of the patients were current smokers at the time of the data collection period, 3.3% were diabetics, 3.3% had hypertension and 3.3% had other respiratory diseases. Likewise, approximately 8.4% of the patients in the study were older than 65 years (Table 1).

The sensitivity and specificity of both AI-based qXR software and radiologist readings were calculated from both sites. The reference test was considered as GeneXpert Test for the analysis. The sensitivity of the test of AI-based qXR software was 100%, (95% CI: 40 - 100%) and specificity 80% (95% CI: 73 - 87%) (Table 2). Similarly, the positive predictive value was 13% (95% CI: 3 - 31%) and the negative predictive value was 100% (95% CI: 96 - 100%). Similarly, the sensitivity of the test by radiologist was 100%, (95% CI: 40 - 100%); and specificity was 62% (95% CI: 53 - 70%). Likewise, the positive predictive value was 8% (95% CI: 2 - 18%) and the negative predictive value was 100% (95% CI: 96 - 100%) (Table 3).

Flowchart: Study Design

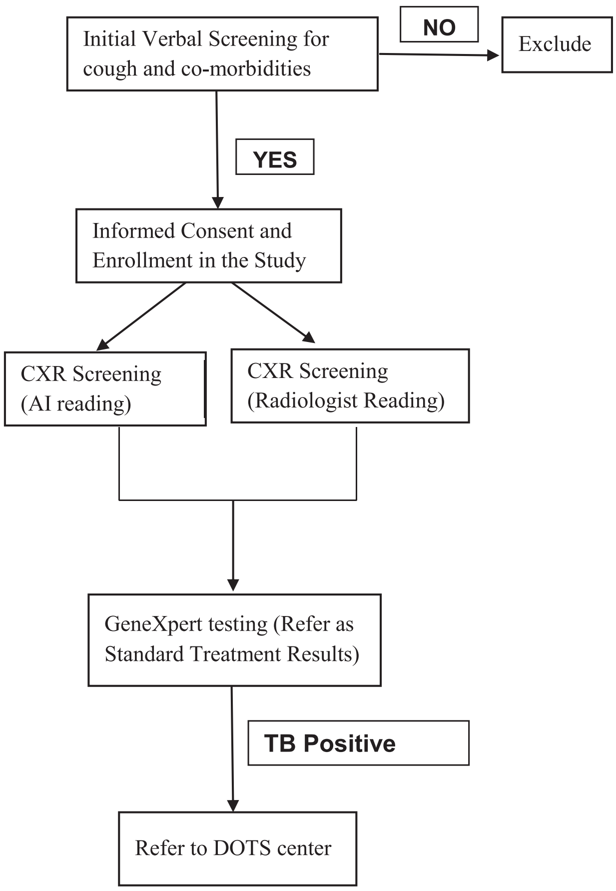

Table 1. Socio-demographics and comorbidity data of patients.

Variables		Frequency (%)
Gender	Male	84 (50%)
	Female	84 (50%)
Comorbidities	Smoker	37 (11%)
	Diabetes	11 (3.3%)
	Hypertension	15 (4.5%)
	Cancer	1 (0.3%)
	Other respiratory conditions	11 (3.3%)
	Age >65 years	28 (8.4%)

Table 2. 2x2 Contingency Table for calculating the Sensitivity and Specificity of AI-Based qXR Software Readings.

AI-based qXR Software	GeneXpert Test		
	Positive	Negative	Total
TB Presumptive	4	25	29
TB Negative	0	104	104
Total	4	129	133

Sensitivity: 100%, (95% CI: 40-100%) and Specificity:80% (95% CI: 73-87%); Positive predictive value - 13% (95% CI: 3 - 31%) and Negative predictive value 100% (95% CI: 96 - 100%).

Table 3. 2x2 Contingency Table for calculating the Sensitivity and Specificity of Radiologist Readings.

Radiologist	**GeneXpert Test**		
	Positive	Negative	Total
TB presumptive	4	49	53
TB Negative	0	80	80
Total	4	129	133

Sensitivity: 100%, (95% CI: 40-100%) and Specificity:62% (95% CI: 53-70%); Positive predictive value - 8%(95% CI: 2 - 18%) and negative predictive value 100% (95% CI: 96 - 100%)

DISCUSSION

The study observed the performance of AI-enabled CXR as a screening tool in TB case diagnosis at two tertiary hospitals in Nepal and reported higher sensitivity for both AI-enabled CXR and radiologist reading. However, the specificity of the radiologist reading in this study was lower than that of the AI-based CXR reading. A plausible explanation for this phenomenon could be attributed to two primary factors. Firstly, despite the radiologists being blinded, their awareness of the study objectives may have inadvertently introduced bias, potentially leading them to over-diagnose TB cases. Secondly, the data was collected during the third phase of the COVID-19 pandemic in Nepal, a period marked by a high volume of patient screenings. This increased workload could have led to radiologist fatigue and, consequently, the overdiagnosis of cases[12] which need to be studied in detail in the future.

To validate the study results, it is crucial to examine other studies conducted in similar settings. A similar study was conducted in India to evaluate the diagnostic accuracy of qXR software using microbiologically confirmed PTB as the reference standard.[13] The study reported a sensitivity of 71% (95% CI: 66%, 76%) and a specificity of 80% (95% CI: 77%, 83%), respectively for qXR software. Whereas, the sensitivity and specificity of radiologists for the detection of microbiologically-confirmed PTB were 56% (95% CI: 50%, 62%) and 80% (95% CI: 77%, 83%), respectively[13] Although the study results are contrary to the study findings, it support the existing evidence that AI-based software's efficiency in diagnosing PTB in medical imaging, with pooled sensitivities of 94% (95% CI 89%–96%) and pooled specificities of 95% (95% CI 91%–97%) for similar developmental studies.[14]

A study in Bangladesh assessed five commercial AI algorithms, including qXR, for triaging tuberculosis using a dataset of 23,954 chest X-rays. The study compared the performance of these AI algorithms against radiologists in detecting TB-suggestive abnormalities in individuals who tested positive for TB using GeneXpert. To facilitate comparison, the researchers created three binary human reading classifications (A–C) by dichotomizing the categories used by radiologists, which varied the radiologist-defined abnormal chest X-ray criteria. When the AI algorithms were adjusted to have the same sensitivity as the radiologists, they demonstrated significantly higher specificity across all three binary classifications compared to the radiologists. Furthermore, a previous study has shown greater accuracy than 95% for detection of TB using AI in chest radiographs.[15]

Similarly, another study assessed the accuracy of radiologists with and without the assistance of a deep-learning model which was trained on chest images from five data sets from Australia, Europe and the USA. The findings reported that radiologists assisted by the deep-learning model achieved higher accuracy than unassisted radiologists for 80% of chest x-ray findings and non-inferior for 95% of findings.[16] Given that humans are susceptible to cognitive biases, such as prejudice and fatigue, which can impair decision-making, AI has the potential to reduce these biases and enhance accuracy in patient care,17 making it a valuable tool to assist radiologists in their readings.

Furthermore, AI-based software has been used to reduce the number of follow-on tests while keeping sensitivity high, providing cost savings that could be applied toward proposed equipment and introduction costs of a Deep Learning system.[18] A previous study mentioned the interpretation of CXR using AI, could reduce the number of GeneXpert MTB/RIF tests needed by 66% while maintaining sensitivity at 95% or better, which possibly could be considered by TB programs where human resources are constrained.[18] Another study further reported reading CXRs with CAD Computer-aided detection software could help reduce the number of higher-priced molecular tests required to confirm diagnosis.[19] In a study conducted in Pakistan, the incremental cost of four different CAD software programs to those of human interpretation of chest X-rays for detecting TB-related abnormalities were compared.[20] The results demonstrated that the cost per screen using CAD software with a perpetual license was significantly lower than the cost of radiologist interpretation for both high-volume ACF and facility-based X-ray testing. This suggests that CAD software could facilitate large-scale screening programs in high TB-burden countries at a lower cost than relying solely on radiologists.

In the Nepalese context, the healthcare system is inequitably weak and fragile and suffers a dearth of trained radiologists in the outskirts and peripheral health facilities. The spread and the use of AI-based CXR throughout the country could potentially reduce the burden on the existing diagnostic testing and confirmatory mechanisms contributing to rapid screening and triaging of presumptive TB cases. Especially in the high burden contexts like ours, the dependency on the passive case-finding approach could be shifted, and more proactive screening could be strengthened at the local level via incorporating new screening technology, reducing the burden of skilled manpower, particularly on the outskirts of the country.[21]

Nepal experienced the third wave of the COVID-19 pandemic during the majority of the data collection period, leading to a huge loss of patients in the study and a higher refusal rate. In addition, due to the fear related to COVID-19 infections during the COVID-19 pandemic, patients were reluctant to provide a sputum sample and screen for CXR, ultimately reducing the number of samples. Similarly, hospital staff and medical students in the study sites were overloaded during the third wave of the COVID-19 pandemic and mostly refused to participate in and perform diagnostic tests.

The smaller sample size of the study could be a major limitation of the study that could impact the accuracy and precision of the results. The wide confidence interval reported indicates the study may be insufficiently powered due to a smaller sample size.[22] The study recommends a formal power calculation to be conducted to determine an appropriate sample size for a full-scale study. In addition, there could be inter-rater variability in the interpretation of X-ray images among radiologists affecting under or over-reporting of cases. However, appropriate training on the Standard Operating Procedure of CXR readings was provided to the radiologist to minimize the discrepancies during formative studies.

Since the study was conducted in two major hospitals across two provinces, the data could be representative of cases of tertiary hospital settings. However, given the data was collected during the COVID-19 pandemic and the lesser number of sample size, the results should be cautiously interpreted.

CONCLUSIONS

Considering, the higher sensitivity and specificity observed for AI readings for the diagnosis of presumptive TB, the spread of deep learning technologies (AI-based CXR software) could potentially reduce the burden on the existing diagnostic test, particularly in settings with less skilled radiologists - contributing to rapid screening and triaging of presumptive TB cases. However, further scientific studies will be required with a larger sample size to verify the findings of the study. Similar pilot programs will be crucial for integrating AI-based software into existing TB screening and diagnostic systems in Nepal. These programs should assess both the effectiveness and cost-efficiency of the software within the local context. Additionally, it is important to evaluate the feasibility and adequacy of physical infrastructure and technical manpower for implementing AI-based software in Nepal.

ACKNOWLEDGEMENTS

We would like to acknowledge International Organisation for Migration, Nepal for the support in conducting this study. A special thanks to Dr. Vasil Gajdadziev, Dr. Biraj Karmacharya, and Dr. Radheshyam Krishna KC for their encouragement in conducting the study. We would also like to acknowledge the effort of Ms. Saleena Shahi, Mr. Samip Khatri, Ms. Suban Thapa for the implementation of the project. A special thanks to Mr. Surendra Madhup, Dr. Mahesh Gautam, Dr. Aasish Singh for your supervision during the data collection period. We would also like to thank Dr. Dipesh Tamrakar, Dr. Raja Ram Dhungana, Mr. Om Prakash Poudel, Dr. Nikesh Mani Shrestha and Dr. Nabin Amatya for the valuable suggestions during development of project. Lastly, we would like to acknowledge Mr. Shirshak Shahi, Mr. Manish Rajbansi, Ms. Prasamsha Bhandari and Ms. Kusum Shilpakar for their enthusiastic participation and tireless support throughout the research study. Lastly, we would like to thank all patients, staff and research team for their participation and support throughout the implementation of the study.

CONFLICT OF INTEREST

There are no conflicts of interest.

REFERENCE

1. NTPS-Report-Bodypages.pdf [Internet]. Available from: https://nepalntp.gov.np/wp-content/uploads/2021/03/NTPS-Report-Bodypages.pdf

2. NTP-Annual-Report-2075-76-2018-19.pdf [Internet]. Available from: https://nepalntp.gov.np/wp-content/uploads/2020/04/NTP-Annual-Report-2075-76-2018-19.pdf

3. Tuberculosis (TB) [Internet]. Available from: https://www.who.int/news-room/fact-sheets/detail/tuberculosis

4. Implementing the End TB Strategy [Internet]. Available from: https://www.who.int/westernpacific/activities/implementing-the-end-tb-strategy

5. Ho J, Fox GJ, Marais BJ. Passive case finding for tuberculosis is not enough. Int J Mycobacteriol. 2016 Dec;5(4):374–8.doi: https://doi.org/10.1016/j.ijmyco.2016.09.023

6. Doshi R, Falzon D, Thomas BV, Temesgen Z, Sadasivan L, Migliori GB, et al. Tuberculosis control, and the where and why of artificial intelligence. ERJ Open Res. 2017 Apr;3(2):00056–2017.doi: https://doi.org/10.1183/23120541.00056-2017

7. Golub JE, Mohan CI, Comstock GW, Chaisson RE. Active case finding of tuberculosis: historical perspective and future prospects. Int J Tuberc Lung Dis. 2005 Nov;9(11):1183–203.[Article]

8. Adhikari N, Joshi LR, Subedi B, Acharya D, Adhikari M, Thapa P, et al. Tuberculosis in Nepal: Situation, Challenges and Ways Forward. SAARC Journal of Tuberculosis, Lung Diseases and HIV/AIDS. 2019 Jul 26;17(1):34–40.doi: https://doi.org/10.3126/saarctb.v17i1.25026

9. qXR AI enabled comprehensive Chest X-ray reporting tool [Internet]. Available from: https://www.qure.ai/product/qxr

10. Du J, Huang Z, Luo Q, Xiong G, Xu X, Li W, et al. Rapid diagnosis of pleural tuberculosis by Xpert MTB/RIF assay using pleural biopsy and pleural fluid specimens. J Res Med Sci. 2015 Jan;20(1):26–31.[FullText]

11. Xpert MTB/RIF implementation manual: technical and operational 'how-to'; practical considerations [Internet]. Available from: https://www.who.int/publications/i/item/9789241506700

12. Gabelloni M, Faggioni L, Fusco R, De Muzio F, Danti G, Grassi F, et al. Exploring Radiologists' Burnout in the COVID-19 Era: A Narrative Review. Int J Environ Res Public Health. 2023 Feb 14;20(4):3350.[Article]

13. Nash M, Kadavigere R, Andrade J, Sukumar CA, Chawla K, Shenoy VP, et al. Deep learning, computer-aided radiography reading for tuberculosis: a diagnostic accuracy study from a tertiary hospital in India. Sci Rep. 2020 Jan 14;10(1):210.[Article]

14. Zhan Y, Wang Y, Zhang W, Ying B, Wang C. Diagnostic Accuracy of the Artificial Intelligence Methods in Medical Imaging for Pulmonary Tuberculosis: A Systematic Review and Meta-Analysis. J Clin Med. 2022 Dec 30;12(1):303.doi: https://doi.org/10.3390/jcm12010303

15. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology. 2017 Aug;284(2):574–82.doi: https://doi.org/10.1148/radiol.2017162326

16. Seah JCY, Tang CHM, Buchlak QD, Holt XG, Wardman JB, Aimoldin A, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. The Lancet Digital Health. 2021 Aug 1;3(8):e496–506.doi: https://doi.org/10.1016/S2589-7500(21)00106-0

17. Khalifa M, Albadawy M. AI in diagnostic imaging: Revolutionising accuracy and efficiency. Computer Methods and Programs in Biomedicine Update. 2024 Jan 1;5:100146.doi: https://doi.org/10.1016/j.cmpbup.2024.100146

18. Qin ZZ, Sander MS, Rai B, Titahong CN, Sudrungrot S, Laah SN, et al. Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems. Sci Rep. 2019 Oct 18;9(1):15000.doi: https://doi.org/10.1038/s41598-019-51503-3

19. Philipsen RHHM, Sánchez CI, Maduskar P, Melendez J, Peters-Bax L, Peter JG, et al. Automated chest-radiography as a triage for Xpert testing in resource-constrained settings: a prospective study of diagnostic

accuracy and costs. Sci Rep. 2015 Jul 27;5(1):12215. doi: https://doi.org/10.1038/srep12215

20. Bashir S, Kik SV, Ruhwald M, Khan A, Tariq M, Hussain H, et al. Economic analysis of different throughput scenarios and implementation strategies of computer-aided detection software as a screening and triage test for pulmonary TB. PLoS One. 2022 Dec 30;17(12):e0277393.doi: https://doi.org/10.1371/journal.pone.0277393

21. Sharma A, Sharma A, Malhotra R, Singh P, Chakraborty RK, Mahajan S, et al. An accurate artificial intelligence system for the detection of pulmonary and extra pulmonary Tuberculosis. Tuberculosis. 2021 Dec 1;131:102143.doi: https://doi.org/10.1016/j.tube.2021.102143

22. Nayak BK. Understanding the relevance of sample size calculation. Indian J Ophthalmol. 2010;58(6):469–70. doi: https://doi.org/10.4103/0301-4738.71673