Development and evaluation of machine-learning methods in whole-body magnetic resonance imaging with diffusion weighted imaging for staging of patients with cancer: the MALIBO diagnostic test accuracy study

Andrea Rockall,[1,7*] Xingfeng Li,[1] Nicholas Johnson,[2] Ioannis Lavdas,[1] Shalini Santhakumaran,[2,3] A Toby Prevost,[2] Dow-Mu Koh,[4] Shonit Punwani,[5] Vicky Goh,[6] Nishat Bharwani,[1,7] Amandeep Sandhu,[7] Harbir Sidhu,[5,8] Andrew Plumb,[5] James Burn,[7] Aisling Fagan,[7] Alf Oliver,[5] Georg J Wengert,[1,10] Daniel Rueckert,[9] Eric Aboagye,[1] Stuart A Taylor,[5,8] Ben Glocker[9] and The MALIBO Investigators

[1]Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK
[2]Nightingale-Saunders Clinical Trials and Epidemiology Unit, King's College London, London, UK
[3]King's Cancer Prevention Group, School of Cancer and Pharmaceutical Sciences, King's College, London, UK
[4]Royal Marsden Hospital and The Institute of Cancer Research, Sutton, UK
[5]Centre for Medical Imaging, University College London, London, UK
[6]Cancer Imaging, School of Biomedical Engineering and Imaging Sciences, King's College London and Department of Radiology, Guy's and St Thomas' Hospitals NHS Foundation Trust, London, UK
[7]Imaging Department, Imperial College Healthcare NHS Trust, London, UK
[8]Department of Radiology, University College London Hospital, London, UK
[9]Faculty of Engineering, Department of Computing, Imperial College London, London, UK
[10]Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, Vienna General Hospital, Vienna, Austria

*Corresponding author a.rockall@imperial.ac.uk

Published October 2024
DOI: 10.3310/KPWQ4208

Plain language summary

Development and evaluation of machine-learning methods in whole-body magnetic resonance imaging with diffusion weighted imaging for staging of patients with cancer: the MALIBO diagnostic test accuracy study

Efficacy and Mechanism Evaluation 2024; Vol. 11: No. 15
DOI: 10.3310/KPWQ4208

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Plain language summary

Whole-body magnetic resonance imaging demonstrates the entire body and can detect the spread of tumour, without the burden of ionising radiation. Recently, the STREAMLINE study reported that whole-body magnetic resonance imaging is accurate, efficient and cost-effective for cancer staging. However, whole-body magnetic resonance imaging is complex to report.

Machine learning is a type of artificial intelligence whereby a computer learns from being given previous data to undertake a task, using techniques such as classification forests, convolutional neural networks, and multi-atlas approaches. Our aim was to develop a machine-learning method to automatically detect lesions on whole-body magnetic resonance imaging to support radiologists by potentially improving their ability to correctly detect disease and reduce the reading time of whole-body magnetic resonance imaging scans in patients with cancer.

Firstly, whole-body magnetic resonance imaging scans from 51 healthy volunteers were used to develop machine-learning methods to automatically detect normal organs.

Secondly, machine-learning methods were trained to detect cancer lesions, using 271 whole-body magnetic resonance imaging scans from a previous study.

Finally, the refined machine-learning technique was tested in 188 patient scans from a previous study, to see if the technique could improve radiology reporting by increasing accuracy and speed in detecting disease. We designed a system to test the accuracy of radiologists reading whole-body magnetic resonance imaging with or without machine-learning support in a near-real clinical National Health Service setting. Twenty-five independent radiologists (18 experienced in reading whole-body magnetic resonance imaging and 7 radiologists inexperienced in whole-body magnetic resonance imaging) were randomly allocated whole-body magnetic resonance imaging scans to read with or without machine-learning support. We found that machine-learning support resulted in similar accuracy for detecting disease and was slightly more efficient in the reading time than for radiological reads without machine-learning support. Differences in interpretation between experienced readers were considered moderate in both cases.

Overall, the study was an ambitious attempt to undertake a highly complex machine-learning task, to detect cancer on whole-body magnetic resonance imaging. Many important steps have been taken but the current machine-learning algorithm did not result in a significant improvement in the radiologist's accuracy for disease detection, although it may have slightly reduced the time taken to read the study. Future work is advocated to further develop machine-learning tools to improve the accuracy of tumour detection.

Efficacy and Mechanism Evaluation

ISSN 2050-4373 (Online)

A list of Journals Library editors can be found on the NIHR Journals Library website

Efficacy and Mechanism Evaluation (EME) was launched in 2014 and is indexed by Europe PMC, DOAJ, Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and NCBI Bookshelf.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full EME archive is freely available to view online at www.journalslibrary.nihr.ac.uk/eme.

Criteria for inclusion in the *Efficacy and Mechanism Evaluation* journal
Manuscripts are published in *Efficacy and Mechanism Evaluation* (EME) if (1) they have resulted from work for the EME programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

EME programme

The Efficacy and Mechanism Evaluation (EME) programme funds ambitious studies evaluating interventions that have the potential to make a step-change in the promotion of health, treatment of disease and improvement of rehabilitation or long-term care. Within these studies, EME supports research to improve the understanding of the mechanisms of both diseases and treatments.

The programme supports translational research into a wide range of new or repurposed interventions. These may include diagnostic or prognostic tests and decision-making tools, therapeutics or psychological treatments, medical devices, and public health initiatives delivered in the NHS.

The EME programme supports clinical trials and studies with other robust designs, which test the efficacy of interventions, and which may use clinical or well-validated surrogate outcomes. It only supports studies in humans and where there is adequate proof of concept. The programme encourages hypothesis-driven mechanistic studies, integrated within the efficacy study, that explore the mechanisms of action of the intervention or the disease, the cause of differing responses, or improve the understanding of adverse effects. It funds similar mechanistic studies linked to studies funded by any NIHR programme.

The EME programme is funded by the Medical Research Council (MRC) and the National Institute for Health and Care Research (NIHR), with contributions from the Chief Scientist Office (CSO) in Scotland and National Institute for Social Care and Health Research (NISCHR) in Wales and the Health and Social Care Research and Development (HSC R&D), Public Health Agency in Northern Ireland.

This article

The research reported in this issue of the journal was funded by the EME programme as award number 13/122/01. The contractual start date was in February 2015. The draft manuscript began editorial review in October 2020 and was accepted for publication in July 2023. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The EME editors and production house have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

This article presents independent research. The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the MRC, the EME programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the EME programme or the Department of Health and Social Care.

This article was published based on current knowledge at the time and date of publication. NIHR is committed to being inclusive and will continually monitor best practice and guidance in relation to terminology and language to ensure that we remain relevant to our stakeholders.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).