

Mammography Reading with Computer-Aided Detection (CAD): Performance of Different Readers

Susan M. Astley¹, Stephen W. Duffy², Caroline R.M. Boggis³, Mary Wilson³, Nicky B. Barr³, Ursula M. Beetles³, Miriam A. Griffiths³, Anil Jain³, Jill Johnson³, Rita M. Roberts³, Heather Deans⁴, Karen Duncan⁴, Geeta Iyengar⁴, Olorunsola Agbaje², Pamela M. Griffiths¹, Magnus A. McGee⁵, Maureen G.C. Gillan⁶, and Fiona J. Gilbert⁶

¹ Department of Imaging Science & Biomedical Engineering, University of Manchester, Manchester, UK

{Astley, Griffiths, sue.astley}@manchester.ac.uk

² Department of Epidemiology, Mathematics & Statistics, Wolfson Institute of Preventive Medicine, London, UK

{Duffy, stephen.duffy}@cancer.org.uk,

{Agbaje, Olorunsola.Agbaje}@cancer.org.uk

³ Nightingale Centre, Withington Hospital, Manchester, UK

{Boggis, Wilson, Barr, Beetles, Griffiths, Jain, Johnson, Roberts, caroline.boggis}@dsl.pipex.com

⁴ NE Scotland Breast Screening Centre, Aberdeen UK

{Deans, Duncan, Iyengar, heather.deans}@nhs.net

⁵ Department of Public Health & General Practice, Christchurch School of Medicine, NZ

{McGee, magnus.mcgee}@chmeds.ac.nz

⁶ Department of Radiology, University of Aberdeen, UK

{Gilbert, Gillan, f.j.gilbert}@abdn.ac.uk

Abstract. Computer-aided detection (CAD) systems place prompts in digital images to attract readers' attention to potential malignancies. A reader must then decide whether or not prompted regions correspond to genuine abnormalities and has the option of disregarding falsely prompted regions. In this paper we investigate different readers' performance with CAD in the context of breast screening. In a retrospective study, eight consultant radiologists each read over 1000 screening mammograms comprising normal cases, screen detected cancer cases and cases that were detected as cancers subsequently. We present their results in terms of cancer detection and recall rates, and relate this to their previous experience of film reading. Our results show that the detection of cancers did not differ significantly between readers, although more experienced film readers were less likely to recommend that normal cases should be recalled.

1 Introduction

Computer-aided detection (CAD) systems can be used to prompt abnormalities that might otherwise be overlooked by a film reader [1]. In order to achieve sufficient sensitivity to genuine abnormalities, CAD algorithms also prompt normal regions

which have some abnormal features [2]. It is the responsibility of the reader using the CAD system to decide whether prompted regions are actually abnormal, or whether the prompts should be disregarded.

Readers in the UK Breast Screening Programme are closely monitored and subject to a number of regulations to ensure good performance; readers must read more than 5,000 screening mammograms per year, participate in clinical meetings and assessment clinics, and have their performance evaluated on a set of difficult mammograms. Despite this, there is still variation in performance between readers. Some of this may be due to experience; the most experienced readers in the breast screening programme have read many tens of thousands of films, whilst newly eligible readers will have read a much smaller number.

Computer-aided detection has been suggested as a way of improving the performance of readers, but different readers may respond in different ways to CAD prompts [3]. In this paper we use data from a study comparing double reading to single reading with CAD to investigate whether there were significant differences in performance with CAD between the eight consultant radiologists who took part.

2 Methods

In the CAD Evaluation Trial (CADET) [4], 10,096 mammograms, originally double read in 1996, were re-read in 2003 by a single reader using an ImageChecker M1000 CM System (R2 Technology, Inc. Sunnyvale, CA) with software version 5.0, a CheckMate Ultra Display Unit with PeerView and Variable Sized Markers [5]. The case-mix was a cohort from the UK National Health Service Breast Screening Programme (NHSBSP) enriched by 50% by the random removal of normal cases. The cases were randomised in such a way that readers did not re-read in 2003 any cases that they had previously read in 1996.

The eight radiologist readers in the CADET study had between 2 and 15 years experience when the films were single read with CAD, and those who also participated in double reading in 1996 had between 1 and 8 years experience at that time. Other readers also participated in the double reading process, and overall the levels of experience of the readers in 1996 and 2003 were approximately similar.

The eight radiologists were based at two screening centres, each of which had different protocols for recalling cases. When single reading with CAD, these protocols were followed as closely as possible. In one of the screening centres (Aberdeen) a case was recalled when double reading if either reader recommended it. When single reading with CAD, a case was recalled if the single reader recommended it. In the second centre (Manchester) the recall process in double reading involved each of the readers scoring the case, and depending on the maximum score cases were either returned to normal screening, discussed with another reader or recalled. When single reading with CAD, cases were also scored and treated in a similar way. All readers in the CADET study underwent a two month training period in the use of CAD prior to commencing reading [6].

In this analysis we review the cancer detection rates for each reader (single reading the 1996 films with CAD) for those cases which had cancer detected by double reading in 1996, and those for which cancer was detected subsequently as interval