Use of Prompt Magnitude in Computer Aided Detection of Masses in Mammograms

Nico Karssemeijer

Radboud University Nijmegen Medical Centre
Department of Radiology
The Netherlands

Abstract. Systems for computer aided detection of masses may be used more effectively when they are used for interpretation of suspect abnormalities, instead of solely using them as a prompting aid to avoid oversights. To use CAD algorithms for detection of masses as a decision aid it may be helpful to display suspiciousness of regions computed by CAD. In this paper the quality of probabilities computed for masses by a commercial CAD system is studied in two ways: 1) by comparing standalone performance of the system to that of experienced screening radiologists, and 2) by determining results of independent double reading with CAD. The study involves results of 15 readers who each read 500 mammograms, and two releases of the CAD algorithm. Independent double reading results are obtained by combining probabilities of the CAD system with the reader assessment for each localized finding reported by the reader, and by computing the fraction of cancers localized correctly as a function of false positive referrals. It was found that standalone performance of CAD is less than that of any reader in the study. Nevertheless, it was found that performance improves significantly with independent CAD reading, and that use of an improved CAD algorithm lead to significantly better results of the combined reader with CAD.

1 Introduction

In computer aided detection (CAD) systems prompts are displayed on regions identified by a computer as suspicious for breast cancer, after the reader has inspected the mammogram without CAD. These prompts may help radiologists to find cancers that initially were overlooked. Results of some prospective studies confirm that screening results are improved when CAD is used [1],[2]. However, considering the high sensitivity of CAD systems, it is also felt that the technology is less effective than expected, particularly for masses, architectural distortion and asymmetry. The reason for this is that it frequently occurs that radiologists do not act on prompts that later appear to be true positives. This suggest that these cancers were not missed by oversight but misinterpreted. Also in experimental studies evidence is found that the majority of screening errors related to masses may be due to misinterpretation rather than oversight [3], [4].

In a previous study it was found that radiologist may be able to use CAD prompting systems to help with interpretation of masses. The use of the system should be radically different though from what is currently recommended.

Susan M. Astley et al. (Eds.): IWDM 2006, LNCS 4046, pp. 54–60, 2006.

Instead of ignoring prompts on regions already inspected, the reader should reconsider decisions with respect to these regions with the use of CAD. On the other hand, prompts of the computer on regions not identified as potential abnormalities might better be ignored, unless it is a clear abnormality that was overlooked. In this paper this way of using CAD mass prompts explored further. First, a comparison is made of the standalone performance of a CAD system with that of experienced screening radiologists. This comparison shows that probabilities, or suspiciousness levels, of prompted regions computed by the CAD system correlate well with radiologists' findings. Second, we conduct an experiment with two versions of a CAD system, of which the most recent one has higher detection performance. It is investigated if the higher performance of CAD leads to better detection results if probabilities of the CAD system are independently combined with human reader assessments, focusing on areas identified by the readers as potential abnormalities.

By comparing ratings of suspiciousness of CAD and human readers it may be understood better how CAD can assist readers with interpretation. Because of the large number of false positives, in practice readers do not have much confidence in CAD as a decision aid. By showing that performance of a CAD system is in fact close to that of a human reader it may be more easy for a radiologist to recognize it as a system that can truly help with screening.

2 Observer Data and CAD System

In this study we make use of data from an observer study, which has been described in detail previously [3]. Fifteen experienced screening radiologists from different countries were involved in this study of which five can be regarded as leading experts in the area of breast cancer screening. In the study they read 500 mammograms of which 250 were priors of cancer cases. It turned out that in 142 cases a visible lesion could be seen in the prior. In 116 of these cases a mass, architectural distortion or an asymmetry was the major sign. These were selected as the true lesions in this investigation.

In the observer study radiologists had been asked to mark and rate all regions that attracted their attention, also those that they would normally not recall. For their ratings they used a scale of suspiciousness ranging from 0 to 100%. They read the mammograms with priors, as is common in screening. So in fact in the study the priors and former priors of the cancer cases were presented, randomly mixed with the 250 normal cases, which also consisted of two subsequent screenings. In total, the 15 readers marked 7173 findings in the selected sample of cases.

In this study we use mass detection results of the R2 ImageChecker. Two versions of the system were used, with software releases from 2001 and 2004. Each mass prompt of the system had a measure indicating importance of the prompt, which is intended to be used in combination with a threshold to select prompts to be displayed.