# The Impact of Integration of Computer-Aided Detection and Human Observers

Nachiko Uchiyama[1], Noriyuki Moriyama[1], Takayuki Yamada[2], and Noriaki Ohuchi[3]

[1] Department of Cancer Screening, Research Center for Cancer Prevention and Screening, National Cancer Center 5-1-1, Tsukiji, Chuo-Ku, Tokyo, Japan, 104-0045
nuchiyam@ncc.go.jp
[2] Department of Radiology, Tohoku University 1-1, Seiryomachi, Aoba-Ku, Sendai, Miyagi, Japan, 980-8574
[3] Department of Surgical Oncology, Tohoku University 1-1, Seiryomachi, Aoba-Ku, Sendai, Miyagi, Japan, 980-8574

**Abstract.** We evaluated the impact of integration of CAD (Computer-Aided Detection) system and human observers in digital mammography. We com-pared the diagnostic efficacy of non-informed observers and informed observers regarding the CAD system's ability (average FP (false positive) per four images and sensitivity of microcalcifications and mass) to detect cancer. With the informed-group, we previously informed them of the accuracy of CAD. In each group, observers recorded the diagnosis before utilizing the CAD system and after utilizing the CAD system according to BI-RADs category and to six additional categories associated with diagnostic confidence. Regarding diagnostic accuracy, with the informed group, sensitivity and NPV were improved without an increase in FP. On the other hand, the diagnostic accuracy of human observers was influenced by prior notification of CAD's accuracy and by CAD's performance in cancer detection itself.

## 1 Introduction

Recently, the performance of CAD has been improved and CAD is being used clinically in digital mammography[1-5]. In this paper, we evaluate the impact with respect to diagnostic accuracy of integrating CAD and human observers in a clinical environment, specifically that of digital mammography.

## 2 Methods

We utilized an indirect FFDM (full field digital mammography) system: Computed Radiography (CR) system (FCR 5000MA Plus: FUJIFILM, Japan) with 50 microns and non-commercial CAD developed by FUJIFILM, Japan. The CR images were diagnosed utilizing soft-copy reading system. The monitors were LCD (Liquid Crystal Display) with 5M pixels (EIZO NANAO CORPORATION, Japan). The clinical cases in this study were randomly selected from screening mammograms. The total number of cases was 50 including 23 malignant cases ( five cases with masses and microcalcifications, five cases with masses, eight cases with microcalcifications, and five cases with FAD

Susan M. Astley et al. (Eds.): IWDM 2006, LNCS 4046, pp. 105 – 110, 2006.

(focal asymmetric density) or distortion) and 27 benign cases. The number of observers was ten. Three observers were radiologists and seven observers were breast surgeons. All of them were experienced at reading mammograms and passed the examination of reading mammograms in accordance with the committee in charge of the quality control manual for mammography screening in Japan. Their sensitivities and specificities were all over 85.0%. There were two randomly selected groups of observers: non-informed observers and informed observers with regard to the CAD system's accuracy. Before this study, we instructed observers in ten cases utilizing other cases including malignancies and normal cases in which CAD pointed out the lesion. The informed-observers were previously informed of the ability of the CAD system with regard to detection rate in microcalcifications and masses and the number of FP per four images. The non-informed observers were not given information regarding accuracy of CAD which was a follows: the average FP marker rate was 1.6 markers per normal 4-view case, sensitivity in calcification was 100.0% and sensitivity in mass was 71.1%. Observers recorded the diagnosis and the schema before utilizing CAD and after utilizing CAD according to BI-RADs category and to six categories associated with diagnostic confidence of malignancy (definitely malignant: 6, probably malignant: 5, maybe malignant: 4, maybe not malignant: 3, probably not malignant: 2, and definitely not malignant: 1). Categories 1 to 3 were evaluated as benign and 4-6 were evaluated as malignant. Diagnostic accuracy was evaluated with respect to sensitivity, specificity, NPV (negative predictive value), PPV (positive predictive value), and ROC analysis utilizing ROCKIT software (Version 0.9.1 BETA).

## 3 Results

1) Sensitivity, Specificity, PPV, and NPV (Table 1-2.)

**Table 1.** Sensitivity, Specificity, PPV, and NPV with the Non-Informed Group

a) Pre-CAD

| | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|
| Observer1 | 0.739 | 0.974 | 0.895 | 0.967 |
| Observer2 | 0.609 | 1.000 | 1.000 | 0.952 |
| Observer3 | 0.782 | 0.974 | 0.900 | 0.972 |
| Observer4 | 0.696 | 0.987 | 0.941 | 0.962 |
| Observer5 | 0.783 | 1.000 | 1.000 | 0.973 |
| Average | 0.722 | 0.985 | 0.947 | 0.965 |

b) Post-CAD

| | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|
| Observer1 | 0.739 | 0.974 | 0.895 | 0.967 |
| Observer2 | 0.609 | 1.000 | 1.000 | 0.952 |
| Observer3 | 0.782 | 0.974 | 0.900 | 0.972 |
| Observer4 | 0.696 | 0.987 | 0.941 | 0.962 |
| Observer5 | 0.783 | 1.000 | 1.000 | 0.973 |
| Average | 0.722 | 0.985 | 0.947 | 0.965 |