

[論 文]

マンモグラム読影における医師とCADシステムの検出結果の比較検討

篠原 範充, 松原 友子[†], 原 武史^{††},
藤田 広志^{††}, 岩瀬 拓士^{†††}, 遠藤登喜子^{††††}

岐阜大学大学院工学研究科電子情報システム工学専攻・〒501-1193 岐阜市柳戸 1-1

[†]名古屋文理大学情報文化学部情報文化学科・〒492-8520 稲沢市稲沢町前田 365

^{††}岐阜大学大学院医学研究科再生医科学専攻知能イメージ情報分野・〒501-1194 岐阜市柳戸 1-1

^{†††}癌研究会附属病院乳腺外科・〒170-845 東京都豊島区上池袋 1-37-1

^{††††}名古屋医療センター・〒460-0001 名古屋市中区三の丸 4-1-1

(2004年7月20日, 最終2004年8月1日受理)

Comparison and Discussion of Doctors' and CAD System's Detection Results
in Mammogram Interpretation

Norimitsu SHINOHARA, Tomoko MATSUBARA[†], Takeshi HARA^{††},
Hiroshi FUJITA^{††}, Takuji IWASE^{†††} and Tokiko ENDO^{††††}

Electronic and Information Systems Engineering Division, Graduate School of Engineering, Gifu University
1-1 Yanagido, Gifu City, 501-1193, Japan

[†]Department of Information Culture, School of Information Culture, Nagoya BunriUniversity
365 Maeda, Inazawa-cho, Inazawa City, 492-8520, Japan

^{††}Department of Intelligent Image Information, Regeneration and Advanced Medical Sciences
Graduate School of Medicine, Gifu University, 1-1 Yanagido, Gifu City, 501-1194, Japan

^{†††}Department of Breast Surgery, Cancer Institute Hospital
1-37-1 Kami-ikebukuro, Toshimaku, Tokyo, 170-8455, Japan

^{††††}National Hospital Organization Nagoya Medical Center
4-1-1 Sannomaru, Naka-ku, Nagoya City, 460-0001, Japan

(Received July 20, 2004, in final form August 1, 2004)

Abstract : We have been developing a computer-aided diagnosis(CAD)system for mammograms and have reported our evaluation studies. The purpose of this study is to evaluate the potential effectiveness of the system in detecting masses and clustered microcalcifications by comparing 142 doctors' and the CAD system's detection results with 200 mammograms for 100 patients employed in doctors' short training course. The sensitivity in the detection of mass was 85% with 2.1 false positives(FPs)per image and that of clustered microcalcifications was 100% with 0.8 FPs per image for this database. We analyzed the data based on the interpretation experiences of the doctors. We found that our CAD system was able to detect some of the candidates which were missed by some doctors. Especially, the system shows the possibility of the stable assistance in detection for the physicians with less experience of the reading. In conclusion, these results imply that our CAD system has a potential to improve the accuracy of the image interpretation.

Key words : mammogram, computer-aided diagnosis(CAD), short training course in mammogram interpretation

1. はじめに

乳がん罹患率は世界的に増加傾向にあり, この早

期発見のため, 視触診やマンモグラフィによる診断が行われている。欧米諸国では, 信頼性のある画像診断法として, 乳がん検診にマンモグラフィが用い

られている。わが国においても、厚生省(現 厚生労働省)からの通達[1]により、平成13年度から乳がん検診にマンモグラフィを導入する自治体が急増している。しかし、残念ながら乳がんの見落とし率は約30%もあるという報告もある[2]。

マンモグラムにおいて乳がんを疑う重要な所見は、腫瘍陰影、微小石灰化クラスタ、および構築の乱れである。われわれは、コンピュータの解析結果を提示することによって、読影医の補助を目的とするコンピュータ支援診断(Computer-aided Diagnosis: CAD)システムの開発を行い、腫瘍陰影[3-6]と微小石灰化クラスタ[7, 8]、構築の乱れ[9, 10]に関する報告などを行っている。

マンモグラムのCADシステムでは、R2 Technology社のImageCheckerが、1998年に世界で初めて商品化された。これ以降も、いくつかの企業がマンモグラム専用のCADシステムの実用化に成功している[11, 12]。

CADシステムの有効性に関しては、シカゴ大のChanら[13]や国立がんセンター東病院のNawanoら[14]が報告している。われわれも40人の医師[15]と597人の医師[16]の読影結果とCADシステムの検出結果を比較し、その有効性を検討した。しかしながら、これらの比較実験では、対象とする医師の読影経験を考慮しておらず、その違いによるCADの有効性に関しては評価できていない。

そこで本研究では、われわれが開発しているCADシステムの検出性能と、経験別の医師の読影結果を比較し、潜在的なCADシステムの有効性の検証を目的とする。

2. 方法

医師の読影結果とCADシステムの検出性能を比較するために用いたマンモグラムは、平成12年度に行われたマンモグラフィ読影講習会で提示された画像100症例(200枚)で、すべて内外斜位(Mediolateral Oblique: MLO)方向から撮影されている。このうち悪性は48枚であり、その内訳は、腫瘍陰影が34枚、微小石灰化クラスタが10枚、構築

Table 1 The 142 physicians' interpretation experience in terms of the number of films.

Level	No. of reading films	No. of physicians
1	0~200	31
2	201~1000	52
3	1001~	59

の乱れが11枚である。これらは、マンモグラフィ診断に習熟した複数の医師が、マンモグラフィガイドライン[17]に従ってカテゴリ分類をし、半数以上の医師の分類結果が一致した画像である。

この画像については、すでに医師の読影結果が分かっているため、その画像をCADシステムによって解析し、両者の検出結果を比較することとした。この実験では、医師にはCADシステムの結果は提示されていないため、医師が検出する病変とCADシステムが検出する病変の関係が明らかになる。

これらのマンモグラムを読影した医師は、マンモグラフィ読影講習会に参加した142名である。読影経験別のレベルをTable 1に示す。読影経験のあるフィルムの枚数により、レベル1(初心者)からレベル3(経験者)の3つのレベルに分類する。

医師は、提示された画像を読影して、マンモグラフィガイドラインに基づき、左右の乳房に対して別々にカテゴリの判定を行う。カテゴリ分類は、「1: 異常なし」、「2: 良性」、「3: 良性、しかし悪性を否定できず」、「4: 悪性の疑い」、「5: 悪性」からなる5段階である。このうちの1つに判定し、それをマークシートに記載する。

われわれが開発したCADシステムの検出アルゴリズムをFig. 1に示す。まず、対象画像をレーザーキャナ(Konica LD-5500)により、サンプリング間隔0.1mm、12bit階調、濃度レンジ4.0でデジタル化する。前処理として、病変部の検出対象領域となる乳房領域を抽出する。次に、微小石灰化クラスタと腫瘍陰影を個々の手法で検出し、良悪性を鑑別する。最後に、結果をディスプレイ上に表示する。本システムの検出性能をTable 2に示す。微小石灰化クラスタの真陽性率は100%、1画像当たりの偽陽性数は0.8であった。構築の乱れの検出アルゴリズムは現在開発中[9, 10]であるため、腫瘍陰影と構

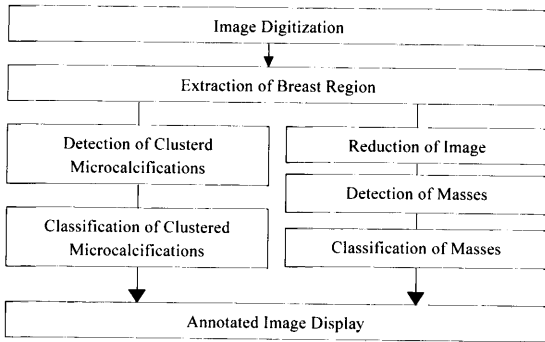


Fig. 1 Flowchart of the overall scheme of CAD system.

築の乱れは、腫瘍陰影の検出アルゴリズムで検出性能を評価した。その真陽性率は73%、1画像当たりの偽陽性数は2.1であり、これまでわれわれが報告してきた腫瘍陰影の真陽性率に比べて低い結果となった。これは、構築の乱れの真陽性率が45%であることが原因である。以下、この結果と医師の検出結果の比較検討を行う。ただし、構築の乱れを検出対象外とした腫瘍陰影の検出結果は、真陽性率85%、1画像当たりの偽陽性数は2.2であった。CADシステムの検出性能は、データベースの画質に大きく影響されるため、高い検出性能を得るためには最適なパラメータを設定する必要がある。それには、データベースの画質は均一であることが望ましい。しかし、今回用いたデータベースは、マンモグラフィ読影講習会用に変更されたものであり、いくつかの施設で撮影された画像から構成されている。そのため、本データベースにおいて、画質のバラツキが確認されたが、今回は実験の性質上、すべての画像に同一のパラメータを用いた。

この読影講習会は、集団検診における医師の検出能の向上を目的としている。よって、医師がカテゴリ3以上の症例を正しくカテゴリ3以上と判断できた場合に、「検出可の症例」とする。一方、カテゴリ2以下の場合には、「検出不可の症例」とする。医師の検出結果とCADシステムの検出結果を、「医師のみが検出可」、「医師とCADがともに検出可」、「CADのみが検出可」、および「医師とCADがともに検出不可」の4つに分類する。この分類を142人の医師に対して行い、読影経験別に

Table 2 Performance of our CAD system for the cases employed in this experiment.

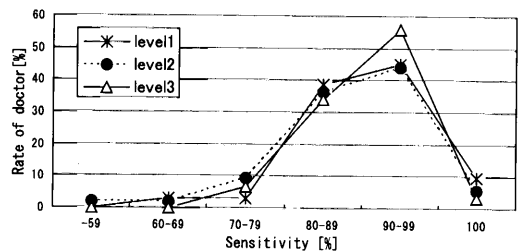
Detection target	Architectural distortion	TPR [%]	No. of FPs / image
Masses	Included	73	2.1
Clustered microcalcifications		100	0.8
Masses	Excluded	85	2.2
Clustered microcalcifications		100	0.8

TPR : True Positive Rate, FP : False Positive

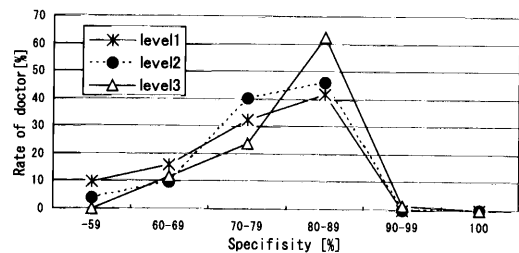
比較検討する。なお、医師とCADの結果比較は、腫瘍陰影、微小石灰化クラスタ、構築の乱れを1つにまとめて行う。

3. 結果

医師の読影経験別(レベル1~3)の感度をFig. 2(a)に、その特異度をFig. 2(b)に示す。感度とは、病変を正しく病変と判断した割合を示す。特異度とは、正常を正しく正常と判断した割合を示す(ここでは、前項で説明したように、画像単位で計算を行った)。つまり、感度と特異度が高い値であれば、読



(a)



(b)

Fig. 2 Doctors' sensitivity (a) and specificity (b) based on their reading experiences in terms of the total number of mammograms.

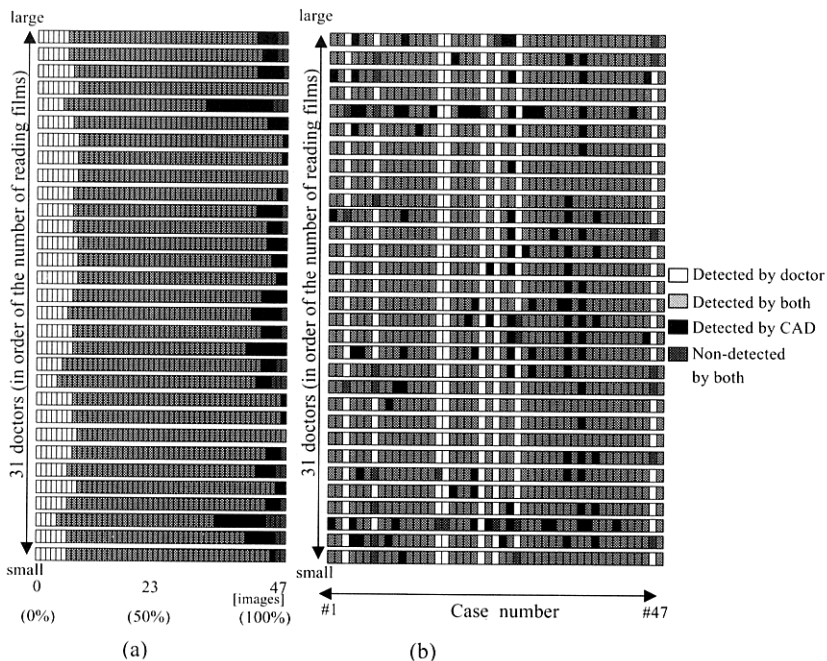


Fig. 3 Comparing 31 doctors' (Level 1) interpretation results and CAD system detection. Distributions based on the number of detected image (a) and cases from #1 to #47 (b).

影能力が高いといえる。ここで、感度、特異度ともに良好といえる90%以上に着目する。感度、特異度ともにレベル1, 2の医師の比率よりも、レベル3の医師の比率の方が大きい。このことから、読影経験枚数が多いほど、読影能力が高い傾向があるといえる。

まず、レベル1の医師31名の読影経験ごとの検出結果とCADシステムの検出性能の比較結果をFig. 3に示す。Fig. 3(a)において、医師の検出は、「医師のみが検出可」と「医師とCADがともに検出可」の二つを合わせた領域となる。CADシステムの検出は、「医師とCADがともに検出可」と「CADのみが検出可」の二つを合わせた領域となる。「CADのみが検出可」の領域が、実際の読影において、CADシステムの補助により医師の読影結果の向上が期待できる領域である。また、「医師とCADがともに検出不可」の領域も存在する。Fig. 3(b)に、異常症例ごとの検出結果を示す。「CADのみが検出可」と「医師とCADがともに検出不可」の領域は、特定の画像に多くみられる。このような画像が医師の

検出率を低下させている。レベル1の医師がCADシステムの検出結果を有効に利用すれば、平均3.8症例(7.9%)の検出率の向上が期待できる。

レベル2の医師52名の検出結果とCADシステムの検出性能の比較結果は、レベル1の医師のそれと同様の傾向を示し、平均4.1症例(8.7%)の検出率の向上が期待できる結果が得られた。また、異常症例ごとの検出性能についても、レベル1の医師と同様、「CADのみが検出可」と「医師とCADがともに検出不可」の領域は特定の画像に多くみられる。

レベル1の医師より読影能力が高いと考えられるレベル3の医師59名に関して同様の解析を行った。Fig. 4(a)は、Fig. 3(a)と同様に、医師の検出結果とCADシステムの検出性能の比較結果を示す。両者を比較すると、レベル3の医師の結果の方が、「医師のみが検出可」と「医師とCADがともに検出可」の領域が多い。しかしながら、レベル3の医師の結果にも「CADのみが検出可」の領域がまだ存在していることが分かる。Fig. 4(b)は、Fig. 3(b)と同様に異常症例ごとの検出性能を示した図である。レ

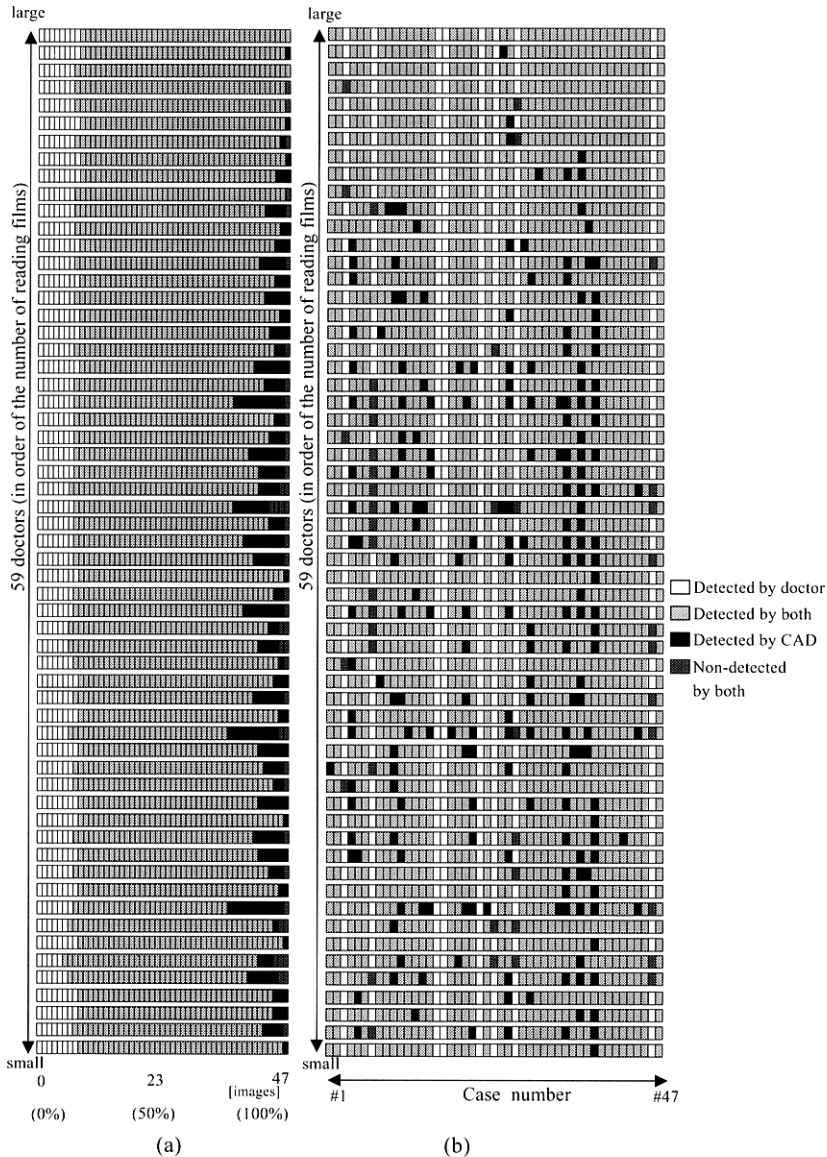


Fig. 4 Comparing 59 doctors' (Level 3) interpretation results and CAD system detection. Distributions based on the number of detected image (a) and cases from #1 to #47 (b).

ベル1の医師と同様、「CADのみが検出可」と「医師とCADがともに検出不可」の領域は、特定の画像に多くみられる。また、CADシステムが検出できない症例のうち「医師のみが検出」の領域が増加している。レベル3の医師についても、CADシステムの検出結果を有効に利用すれば、平均3.8症例(平均8.1%)の検出率向上が期待できる結果が得ら

れた。

次に、レベル1の医師の正解率とその画像の枚数、およびCADシステムの検出枚数とその偽陽性候補の数を Fig. 5 に示す。これにより、医師の正解が低い(検出が困難である)症例に対する、CADシステムの検出状況と偽陽性候補の数が明らかになる。例えば、医師の正解率が70%以上80%未満の症例は

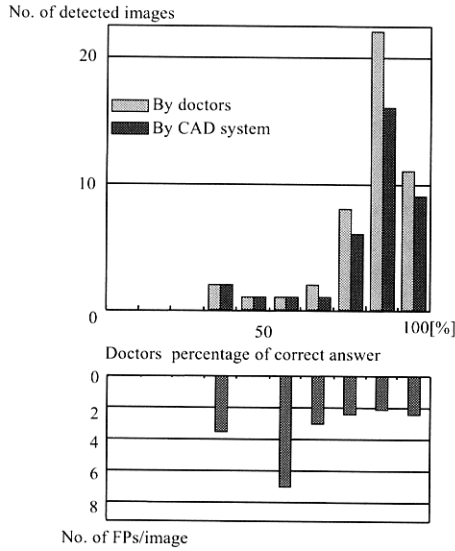


Fig. 5 The relation with the doctors' (Level 1) percentage of correct answer and the performance of our CAD system.

8枚であり、そのうちの6症例をCADシステムが検出することができた。このときの偽陽性数は、1画像当たり2.2個であった。CADシステムは、医師の検出が困難な症例や容易な症例に対しても、一定の検出性能を得ることができる。そのときの偽陽性候補の数も50%以上60%未満の1症例に多く見られるが、それを除けば、医師の正解率によらず一定である。

レベル3の医師の正解率とその画像の枚数、およびCADシステムの検出枚数とその偽陽性候補の数をFig. 6に示す。レベル1と同様に、CADシステムは、医師にとっての読影の難易度に関係なく、一定の検出性能を得ることができる。また、そのときの偽陽性候補の数も、医師の正解率によらず一定である。

医師の見落としが多いがCADシステムが検出した症例をFig. 7に示す。この症例は、142名中93名が指摘できなかった症例であり、腫瘍陰影が厚い乳腺と重なっている。CADシステムは、この症例を正しく指摘することができた。逆に、殆どの医師が指摘したが、CADシステムは検出できなかった症例をFig. 8に示す。これは、142名中140名が指

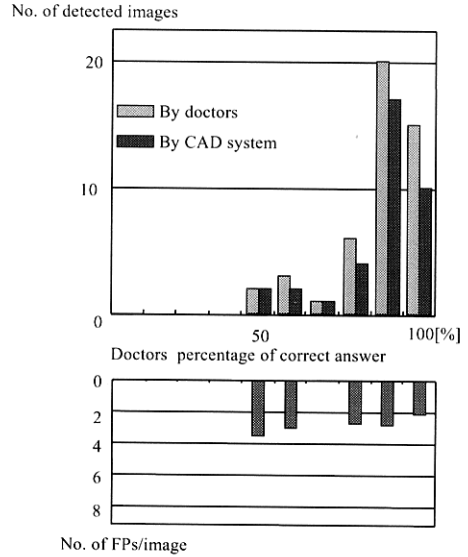


Fig. 6 The relation with the doctors' (Level 3) percentage of correct answer and the performance of our CAD system.

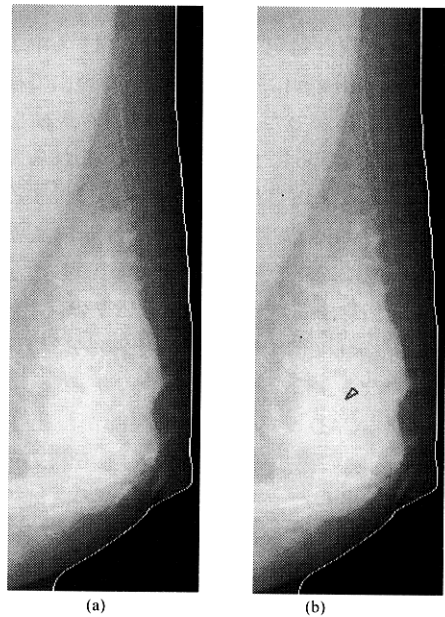


Fig. 7 Images in which 93 of 142 doctors missed the mass. An original image (a) and detected mass by our CAD system (b).

摘した症例である。腫瘍領域は高コントラストであるが、正常な乳腺領域と隣接するため、CADシス

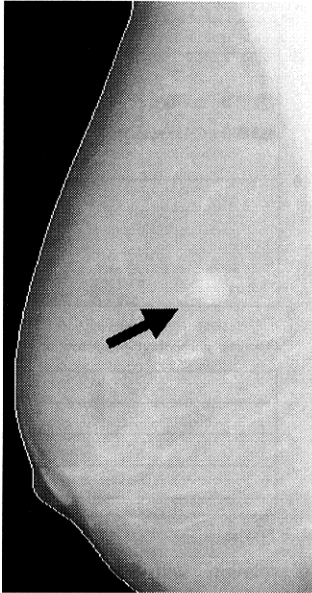


Fig. 8 An image correctly detected by all doctors but missed by our CAD system.

テムは正常構造と判定した。また、142名中138名が指摘した腫瘍陰影に関しても同様の理由で検出することができなかった。CADシステムが検出できなかった症例はさらに6症例あるが、これらの病変はすべて構築の乱れである。この症例に対する医師の検出率は、90%以上が3症例、80%以上が2症例であり、比較的高い検出率であった。しかし、残りの1症例に対する医師の検出率は73%であった。

以上より、それぞれの病変の検出結果についてまとめると次のようになる。

腫瘍陰影：CADシステムは、医師が検出することが困難(医師の感度が60%以下)な6枚のうち5枚を正しく検出した。読影経験枚数別の医師の平均感度は、レベル1で平均89%、レベル2で平均87%、そして、レベル3で平均91%と比較的高い値であった。正答数の標準偏差は、レベル1で4.6、レベル2で4.4、そして、レベル3で3.9であり、読影経験枚数が増えるにつれて安定した検出結果になると言える。また、CADシステムを用いることにより、正答数の標準偏差が、レベル1で1.6、レベル2で2.5、そして、レベル3で3.2に向上する

可能性があることが示された。

構築の乱れ：専用の検出アルゴリズムは、現在開発中であるため、腫瘍陰影の検出アルゴリズムを適応した。CADシステムの検出結果は平均45%であった(開発中の検出アルゴリズムを取り入れることができれば、CADシステムの検出性能の向上が期待できる)。医師の感度は、レベル1で平均89%、レベル2で平均89%、そして、レベル3で平均95%と、すべての読影経験レベルで比較的安定した高い値であった。また、読影経験枚数が増えるにつれて安定した検出結果となった。

微小石灰化クラスタ：CADシステムは、医師が検出することが困難な3枚をすべて正しく検出した。医師の感度は、レベル1で平均97%、レベル2で平均96%、そして、レベル3で平均98%であった。特異度はレベル1で平均88%、レベル2で平均88%、そして、レベル3で平均90%であった。このことから、医師の読影経験枚数による大きな違いは見られず、比較的安定した高い検出結果であると言える。

4. 考 察

マンモグラフィ読影会で提示されたマンモグラムに対し142名の医師の検出結果とCADの検出性能との比較実験を行った。その結果、医師が指摘できない(あるいは、指摘されにくい)症例でもCADシステムが検出できるものが存在することが分かった。また、医師がCADシステムを有効に利用すれば、レベル1の医師で平均3.8症例、レベル2の医師で平均4.1症例、レベル3の医師で平均3.8症例の病変の検出能が向上する。特に病変別で見ると、腫瘍検出の正答数の標準偏差が、レベル1で3.0、レベル2で1.9、そして、レベル3で0.7だけ小さくなり、個々の読影レベルの均一化が期待できる。また、マンモグラフィ読影講習会という性質上、読影経験枚数の差異は少ないと考えられるが、参加した医師の読影経験に関係なく、医師の診断結果を向上させる可能性が示された。特に、読影経験枚数の少ない医師(レベル1と2)に対しては、より安定した高い

読影レベルへの向上が期待できる。

しかし、CADシステムの偽陽性候補の存在により、医師の真陰性率を低下させる可能性があるため、ROC解析などの観察者実験を行い、CADシステムの具体的な臨床応用について検討する必要がある。

5. 結 語

マンモグラフィにおける腫瘍陰影と石灰化クラスタについて、CADシステムの潜在的な有効性を確認することを目的とし、読影枚数に基づいた経験別に分けた医師の検出結果とCADシステムの検出性能との比較実験を行った。その結果、医師が検出することが困難な症例についても、CADシステムは病変を指摘できることが確認された。

本研究では、医師にCADシステムの結果を提示していないため、CADシステムの直接的な有効性は示されていないが、医師の検出結果とCADシステムの検出結果とを合わせることで、検出率は向上し、経験の少ない医師でも高い検出率を得られる可能性が示された。

謝 辞

本研究はマンモグラフィ精度管理中央委員会および多数の医師の協力によるものであり、関係各位に対し、ここに感謝を申し上げます。本研究の一部は、厚生労働省がん研究助成金(遠藤班, 大内班)の助成を受けて行った。

文 献

- [1] 厚生省老人保健福祉局：がん予防重点健康教育及びがん検診実施のための指針とがん検診実施上の留意事項，2000.
- [2] 長谷川 玲：世界で始めて商品化されたマンモグラフィ用CAD—ImageChecker—，日放技誌，56(3)，355-356，2000.
- [3] 松原友子，笠井 聡，関 和泰，他：マンモグラムのためのコンピュータ支援診断システムの開発 - 腫瘍陰影の自動検出における低濃度領域抽出法の改善，日乳癌検診誌，7(1)，87-101，1998.
- [4] 畑中裕司，藤田広志，原 武史，他：乳房X線写真における腫瘍陰影の自動検出 - 一部領域の欠損した腫瘍陰影への対応 - ，医画情誌，17(1)，40-44，2000.
- [5] Hatanaka Y, Hara T, Fujita H, et al.: Development of an automated method for detecting mammographic masses with a partial loss of region, IEEE Trans. Med. Imag., 20(12), 1209-1214, 2001.
- [6] 笠井 聡，松原友子，藤田広志，他：マンモグラム上の乳腺領域周辺に存在する腫瘍陰影検出に特化したアルゴリズムの開発，ME誌，38(2)，111-117，2000.
- [7] 平子賢一，藤田広志，原 武史，他：乳房X線写真における微小石灰化検出フィルタの開発 - 濃度こう配と3重リングフィルタ解析に基づく手法，信学論 D-II, J78-D-II(9)，1334-1345，1995.
- [8] 平子賢一，藤田広志，原 武史，他：コントラスト補正処理と可変リングフィルタ解析を導入した微小石灰化クラスタ検出法，Med. Imag. Tech., 14(6)，665-679，1996.
- [9] 山崎大輔，松原友子，藤田広志，他：乳房X線画像における構築の乱れ領域の自動抽出法，医画情誌，19(2)，69-72，2002.
- [10] 市川徹子，松原友子，原 武史，他：マンモグラムにおける乳腺の集中を伴う構築の乱れ領域の自動検出法，信学論 D-II, J87-D-II(1)，348-352，2004.
- [11] 藤田広志：マンモグラフィCADシステムの現状，Med. Imag. Tech., 21(1)，27-33，2003.
- [12] 特集：CAD最前線，医画情誌，21(1)，2004.
- [13] Chan HP, Doi K, Vyborny CJ, et al.: Improvement in radiologists' detection of clustered microcalcifications on mammograms: The potential of computer-aided diagnosis, Invest. Radiol., 25(10)，1102-1110，1990.

- [14] Nawano S, Murakami K, Moriyama N, et al. : Computer-aided diagnosis in full digital mammography, *Invest. Radiol.*, 34(4), 310-316, 1999.
- [15] 福岡大輔, 原 武史, 遠藤登喜子, 他: 乳房 X線写真における医師の読影とCADシステムの検出結果との比較, *日放技誌*, 56(3), 436-442, 2000.
- [16] 畑中裕司, 松原友子, 原 武史, 他: 医師のマンモグラム読影自習における乳がんとCADシステムの検出結果との関係, *日放技誌*, 58(3), 375-382, 2002.
- [17] マンモグラフィガイドライン委員マンモグラフィガイドライン委員会: マンモグラフィガイドライン. 医学書院, 東京, 1999.