乳癌検診におけるコンピュータ支援診断(CAD) 現状と将来の可能性

シカゴ大学放射線科カートロスマン放射線像研究所

土井 邦雄

Computer-Aided Diagnosis for Screening of Breast Cancer on Mammograms Current Status and Future Potential

Kunio Doi, Ph.D.

Kurt Rossmann Laboratories for Radiologic Image Research

Department of Radiology, the University of Chicago

5841 South Maryland Avenue, Chicago, IL60637

E-mail: k-doi@uchicago.edu

Key words : image processing, computer analysis, breast cancer, mammography, screening

はじめに

コンピュータ支援診断(CAD)は、最近、マモグラフィの分野でかなり多く使われています。推定によると、約4,000台以上の装置が世界中で利用されています。アメリカにおいては、一度、乳癌を検出するためにCADを使った医師の99%の方は、もうこれなしでは読影したくないと言っているとのことです。では、その理由は一体何かということについて、その原因を考えるのに役に立つと思われる材料を以下に述べます。

American Cancer Societyは、1990年に、そのときから25年後には、癌の死亡率を半分に減らすという大きな目標を掲げています。最近、雑誌Cancerの7月号には、その中間報告¹⁾が出ています。それによると、すべての癌については、癌の死亡率は徐々に下がっており、大体1年に1%減少しています。この調子でいくと、公約をした2015年には、大体23%下がる予測が出ています。しか

し、個々の癌に関しては、男性の肺癌、大腸癌、前立腺癌および乳癌はかなり予測された線に近い減少が観察されています。特に乳癌についての死亡率は、年間約2%減少していると報告されています。これはかなり大きなことだと思います。一体なぜこのような減少が起きているかという、乳癌の早期発見と治療法の進歩が大きく寄与していると信じられています。特にその中で、CADの役割はいままではそれほど大きくなかったと思いますが、今後、大きくなることが予測されます。

CADに関する研究²⁻¹⁰⁾を世界的に見ると、多くの研究結果が報告されるのは、RSNA(北米医学放射線学会)です。この学会でのCADに関する発表件数は毎年約200件です。しかし、デジタルマモグラフィ、肺癌集検、CR、DR、フラットパネルディテクターなどに関する全発表件数は100件以下です。そこで、CADに関する研究発表が明らかに多いことがわかります。CADに関する

この原稿は、2006年11月17日(金)、仙台国際センターにおける第16回日本乳癌検診学会総会における特別講演の内容を基に用意されました。

第1図. コンピュータは、マモグラムに含まれる微小石灰化のクラスターを正しく検出しています。この例では偽陽性はありません。

研究発表を臓器別に見ると、最近では、胸部、乳房、大腸の関係が非常に多く、それ以外のものも次々に発表が増えているのが現状です。

世界最初のCADシステムとして構築されたのは、1994年のシカゴ大学のシステムです。現在のCADも、基本的にはこれと大きく変わっていません。CADシステムの構成要素は、フィルム画像をデジタル化するためのレーザースキャナー、デジタル画像を定量的に分析するコンピュータ、その結果を示す表示装置です。多くの画像データは、その当時、光磁気ライブラリーに保存されていました。当時、シカゴ大学では、25,000人のケースについて調査されましたが、左右の乳房と二方向からの撮影があるので、約10万枚以上のマモグラムについて、このコンピュータシステムが応用されました。この研究から、大変興味ある結果が報告されています。乳癌の患者について、乳癌であることが分かったときより1年前、あるいはそれ以前に撮られたマモグラムをコンピュータで調べると、大体その半分(50%)のケースについて

は、コンピュータは早期の癌の病巣を検出していることが分かったのです。したがって、CADを臨床に実用化していたならば、多くの“見落とし”を防ぎ、早期発見ができたのではないかと予想されます。第1図は、コンピュータによる微小石灰化のクラスターの検出結果を示しています。拡大すると、白い点々がつらなっていますが、これが微小石灰化のクラスターの検出されたものです。

一体なぜCADが必要なのかという疑問は、医用画像における病巣の“見落とし”と関係しています。病巣の“見落とし”に関する研究は、過去40～50年間にかなり報告されています¹⁰⁾。また、鑑別診断はかなり難しいともいわれています。特に、胸部写真のノジュールの見落としに関しては、初期の頃からアメリカや日本でも研究されています。その結果、胸部写真を読影する医師は、約30%のノジュールを“見落とし”するというデータが出ています。

一方、乳癌検出のためのマモグラムの検診の始まったのが、いまから30～40年前ですが、アメリ

力で始まった当時は、マモグラムの病巣を“見落とし”することはないといわれていました。その後、いろいろな研究によって、10%から30%は見落としする、さらに生検された疑いのある患者の中でも、実際に乳癌の患者はたったの15%から30%であると報告されています。したがって、乳癌の診断はきわめて難しいということです。また、見落としという観点からは、CT画像は断面像ですので“見落とし”はないといわれたときもありますが、最近の研究によると、CT画像においても、約30%のノジュールは見落としすることが明らかになっています。

では、一体なぜこのような“見落とし”をするのかについて、胸部写真の読影に関しては、人間の目がどのように画像を追跡するかを、ペンシルバニア大学のハル・クンデル¹¹⁾はアイカメラを用いて詳しく研究しています。その結果、病巣を見ていなかったことによる検出エラーが約30%、病巣を見ているのだけれども、病巣であることを認識してないことによる認識エラーが約30%、病巣があるけれども、良性か悪性か、あるいは大事な病巣かどうかという判断になったとき、最終決定で間違えるディシジョンエラーが約40%あると報告されています。これらはいずれもヒューマンエラーです。

アメリカでの乳癌検診においては、左右乳房および二方向からの撮影がされ、患者あたりに4枚のマモグラムが得られます。集検の場合には、約1,000人の中に4人から5人の乳癌患者がいるといわれていますので、250人に1人程度の乳癌しかありません。そこで、年間5,000人を対象にするような検診では、2週間に1人程度の乳癌患者しかいませんから、これを見つけるのはなかなか大変です。大部分の正常な方の画像も、十分注意して見なければならぬことになります。

1. コンピュータ支援診断の現状

そこでコンピュータ支援診断(CAD)とは何かについて、まず定義から始めます。CADは、医用画像に含まれる病巣をコンピュータによって定

量的に分析した結果を第二の意見として利用する医師による診断²⁻¹⁰⁾です。これは自動診断ではありません。コンピュータの結果を利用しますが、最終決定は医師による診断です。これは簡単に思えますが、それほど簡単ではない点があります。例えば、コンピュータの特性とCADの特性は一致するかというと、必ずしも一致するとは限りません。コンピュータによる検出結果がよくても、必ずしもCADの結果はよくなりません。反対にコンピュータの結果が悪くなくても、CADの結果がよくなる場合もあります。そこで、どのようなコンピュータの特性がCADにとって役に立つのかということを考慮する必要があります。例えば、医師が容易に検出できる比較的明らかな病巣だけをコンピュータが検出しても、これは役に立ちません。コンピュータが検出しても、医師が病巣と思えない場合には、これも役に立ちません。ところが、医師がうっかり“見落とし”てしまうような病巣をコンピュータが検出して、これを医師が病巣と認識してアクションがとれる。このような場合には、CADは有用です。したがって、CADの特性をどのように評価するのか、測定するのかというのは、必ずしも簡単ではありません。順番を追って述べます。

まず最初に、多数の病巣を含むケースと正常のケースを含むデータベースを用意します。次は、コンピュータのアルゴリズムを開発して、病巣の検出結果を調べます。この結果を用いて、多くの医師によってコンピュータの結果のない場合とある場合について、病巣を検出するための観察者実験を行います。そして、コンピュータの結果のある場合について、病巣を検出する特性が改善されたかすると、その結果の差に統計的有意差があるかどうかを検討する必要があります。統計的有意差があれば、そのコンピュータの結果は医師にとって役に立つということになります。

次は、コンピュータ支援診断はなぜ必要かという疑問です。すでにお分かりのように、画像診断は主観的判断に基づくものです。したがって、ときどき病巣を見落とししたり、診断のパラツキがあ

ります。最近では、CT などでは非常に多数の画像が得られますので、読影時間の増加が大変困るということです。そこで、コンピュータ支援診断の目的は2つあります。まず1つは、画像診断の品質の向上です。これは正確度と再現性の改良によって実現されます。もう1つは、診断の生産性の向上です。読影時間を短縮するということです。

CAD の一般的手法としては、病巣の位置を見つけること、病気の確率の推定、さらに病巣などの計測と定量化です。乳癌の場合には、乳癌のありそうな場所を見つけることや、良性、悪性の区別などが含まれます。CAD の基本的技術としては3つあります。まず、病巣の強調および抽出をするための像処理です。さらに、病巣と病巣の候補に関しては、特徴量の定量化が必要です。そのあとは、病巣と偽陽性の区別や正常と異常を区別するためのデータ処理、あるいは良性和悪性を区別するためのデータ処理などが含まれます。いままで述べたように、コンピュータ支援診断はきわめて一般的で、幅広い概念ですから、CAD はすべてのモダリティに当てはまります。それらは、X 線写真、CT、MRI、超音波、核医学の画像です。さらに、すべての撮影部位および撮影法、頭部から足の先までが対象になります。また、いろいろな撮影方法にも応用されます。

次に、CAD の現状を簡単に要約します。基礎的および臨床研究の立場からは、本格的な研究の始まったのが、今から約20年前の1983年ごろです。今までのところ、マモグラフィ、胸部単純写真、胸部 CT 画像および腹部大腸 CT 画像などにおいて、病巣の検出あるいは鑑別診断の手法が開発されています。CAD を利用することによって、放射線科医の診断能は明らかに改良されることが、ROC 曲線などを用いて示されています。さらに、臨床実用においては、乳癌検出率は約10%から20%向上することが報告されています。

一方、商品化と実用化の観点からは、1998年にアメリカの R2 Technology というベンチャー企業が、マモグラフィの CAD システムを最初に開発し、FDA の認可を受けています。R2は、現在

Hologic という会社の一部になっています。2001年には、当時、Deus という会社が胸部単純写真のノジュール検出の CAD を開発していますが、その後、会社の名前が Riverain に変わっています。CADx および iCAD という会社もマモグラフィの CAD を開発していますが、その後、合併して iCAD という1つの会社になっています。さらに、Kodak もマモの CAD を開発しています。その後、CT ノジュールに関しては、R2のシステムが開発され、FDA の認可を受けています。さらに、2002年には医療報酬請求の認可がおりて、これが現在の CAD に大きな影響を与えているといわれています。日本国内においては、胸部 CAD、マモ CAD、CT ノジュールのための CAD などがテストされているのが現状です。

マモグラムにおけるコンピュータによる病巣の検出では、かなり複雑なアルゴリズムが使用されていますが、基本的な考え方はきわめて簡単です。例えば、微小石灰化の検出では、デジタルマモグラムに含まれる石灰化の成分を強調する画像処理を行います。その後、石灰化成分を抽出し、それらの特徴量の分析を行い、さらにニューラルネットワークを用いて偽陽性成分を減少させ、最終的に検出された石灰化のクラスターを表示します。**第2図**は、微小石灰化のクラスターの検出感度と偽陽性の数を示します。実線は、1995年頃にシカゴ大学で開発した微小石灰化の検出システムの特性です。これが R2という会社にライセンスされたときの特性です。その後、R2はいろいろな改善を重ねて、左上の四角の点のあたりまで改良されています。これは推定した特性です。これらの結果を比較すると、大学である程度の特性を得られれば、それ以後は企業による集中的な努力によって、かなり大きな特性の改良のできることを示していると思います。今後、いろいろな CAD について、このようなことの実現されることが期待されます。

ROC 曲線は、現在、放射線科医や医師の読影能力を評価したり、モダリティの評価をするのに一番信頼性の高い道具であるといわれています。

第2図. マモグラムにおける微小石灰化クラスターの検出特性の比較です。実線は、1995年ごろのシカゴ大学での研究から得られた特性を示し、左上の黒い四角は、R2社によるCADシステムの最近の特性の推定値です。

ROC 曲線は、病巣を含む多数の画像と多数の正常画像を観察することによって得られる結果について、病巣の正答率を縦軸に、正常画像を“病巣があったかもしれないと間違えて判断する”偽陽性率を横軸にプロットしたものです。曲線になる理由は、判断基準によって正答率と偽陽性率が変化するからです。例えば、判断基準が厳しいときには、原点の近くの左側に移り、判断基準が緩やかになると、右側に移るからです。観察者の特性が高い場合と低い場合を比べると、高い特性は左上のコーナーに位置し、低い特性は右側に下がります。45度の直線は、正答率と偽陽性率が同じになりますので、ランダムに、つまりでたらめに応答したときの特性に対応します。そこで、多くのROC 曲線が得られる場合には、できるだけ左上のほうが優れた特性ということになります。

第3図のグラフは、CAD が役に立つことを初めて証明したデータで、1990年当時、シカゴ大学で研究しており、現在はミシガン大学にいるフェアンピン・チャン¹²⁾が発表したものです。これを見ると、放射線科医が微小石灰化のクラスターを検出する能力は、コンピュータの結果を利用することによって改良されることを示しています。p 値がきわめて低いですから、この改良には統計的

第3図. コンピュータの検出結果を利用しない場合と利用する場合(CAD)における、放射線科医による微小石灰化の検出特性のROC 曲線による比較です。放射線科医の検出特性は、コンピュータの結果を用いることによって改善されています。

有意差があるということになります。先述のように、その当時のコンピュータの検出特性はあまり高くはなく、これが役に立つかどうかは分っていませんでした。しかし、このデータによって、医師はコンピュータの結果を利用できることが証明されたのです。

2. 乳癌検診における CAD の役割と評価

次は、乳癌検診において、乳癌検出に関するCAD の役割の評価について、どうしたらよいか、どのような方法があるかについて述べます。まず最初は、アルゴリズム開発中の評価です。この最初の段階では、少数のマモグラムによる評価でも構わないと思います。その次は、臨床評価が必要です。これは未知の臨床例について、コンピュータだけの特性を調べることです。3 番目は、コンピュータの結果を見て、果たして観察者がこれを利用できるかどうか、つまり、医師とコンピュータの相乗効果があるかどうかを調べる実験です。最終的には、プロスペクティブな臨床研究によって、乳癌を検出する能力が本当に改善されているかどうかを調べます。いろいろなステップがあ

りますので、それぞれについての特性を理解する必要があります。アメリカでは、最後の部分の3つのステップについてかなりの研究報告が出ています。

最初の、コンピュータだけの乳癌の検出率については、カナダのプリティッシュコロンビア大学のリンダ・ワーレン・パーフェン¹³⁾の研究が有名です。2000年に Radiology に報告されています。この人の行った研究は、R2がFDAの認可を取るために集めた、たくさんのケースの分析から始まっています。13の施設から1,000以上の乳癌のケースを集めています。癌が見つかった患者のうち、なるべく近くの過去画像を調べます。そのうち約半分、500程度のケースについては過去画像が見つかっています。その過去画像のうちに、乳癌があると思えるもの、しかも多くの医師が乳癌らしいと判断したためのアクションがとれると考えられるケース、これをアクションナブル・プライア・ミスト・キャンサーと呼んでいます。それらについてもコンピュータの特性を調べています。これらのケースについて、乳癌の検出率は、現在画像の1,083例の癌については84%、“見落とし”と思われる過去画像の115例の乳癌については77%という結果が報告されています。そこで、かなり高い確率で乳癌を見つけることができるということがわかります。

同じケースについて、スタンフォード大学のバードウェル¹⁴⁾が、さらに詳しくアクションナブル・プライア・ミスト・キャンサーについて調べています。これらを、35例の微小石灰化と80例のマス(腫瘍)に分けると、微小石灰化の検出率が86%、マスが73%と、明らかにマスのほうが検出が難しいことを示しています。その後、iCADによる別のCADシステムを使って、ジョージタウン大学のブレム¹⁵⁾が、123例の乳癌について、同じような研究を行っています。ここでの検出率は65%で若干低いのですが、それでも多くの見落としの癌を見つけることができる可能性を示しています。

次は、テキサスのペーラー大学のエバンス¹⁶⁾は、ある特定の種類の乳癌、インベシブ・ロビ

ュラー・カルシノーマだけを94例集めて調べたところ、コンピュータによって91%検出できることが報告されています。一方、デューク大学のジェイ・ベーカー¹⁷⁾の2003年の報告によると、80例のアーキテクチャル・ディストーションについては、その当時のCADのシステムはきわめて感度が低く、検出度は38%しかないという報告が出ています。次は、ドイツのイエナのフリードリッヒ・シラー大学のマリク¹⁸⁾の研究ですが、150例のマスと114例の微小石灰化についての検出を調べた結果、90%および91%と大体同じくらいの検出感度が得られたという報告もあります。その後、ニューヨーク州のバッファローにあるウェンディ・ローガン・クリニックのツーラ・デストーニス¹⁹⁾による52例のアクションナブル・プライア・ミスト・キャンサーについての研究では、これも同様の検出率(71%)が報告されています。

スタンフォード大学のデボラ・イケダ²⁰⁾の研究は、2004年に Radiology に報告されています。この研究では、前年に何かがあったのは認識されていたけれども、特定のはっきりした病巣ではないと判断されたため、明らかに見落としの乳癌の病巣を対象にしています。これらの137件の例を調べると、30%のものはコンピュータが検出していたことを報告しています。次は、カリフォルニアのカイザーパーマナント病院のバトラー²¹⁾の研究では、212の乳癌のうち、違う場所に、臨床的には疑われていなかった場所に、コンピュータによって87%の癌を見つけたという報告があります。デューク大学のソウ²²⁾の研究によると、アモルフオス・カルシフィケーションは微小石灰化の中でも非常に細かい難しい病巣ですが、146例のうちコンピュータは40%しか検出していないという報告があります。さらに、ジョージワシントン大学のブレム^{23, 24)}は、2005年に2つの論文をAJRに報告しています。1つは、自分たちの病院の比較的少ない数のケースについての報告で、もう1つは、18のマルチセンタートライアル²⁴⁾によって、909例の乳癌についてのコンピュータによる検出を調べています。かなり高い検出感度(89%)

が報告されていますが、特に注目されるのは、デンスプレストです。401例のデンスプレストについても88%という高い検出力が得られていることが報告されています。

以上はコンピュータだけの検出特性ですが、次は、人間がコンピュータの結果を利用するとどうなるかに関する研究です。ここでは、3つの条件が比較の対象になります。それらは、1人の医師による読影結果、2人の医師によるダブルリーディング(二重読影)の結果、そして1人の医師によるCADの結果です。一般には、1人の医師によるCADの結果は、1人の医師による結果よりも優れていると期待されていますが、さらにダブルリーディングの結果と同等になれば、ダブルリーディングに必要な2人の医師を、1人に置きかえられる可能性があります。そこで、乳癌の検診にダブルリーディングを用いているイギリスなどでは、マンパワーの節約にとってきわめて重要です。

イタリアのグループの報告²⁵⁾では、7つの大学の医師が集まって、10人の放射線科医によって17の乳癌について読影実験をしています。その結果、ダブルリーディングの場合(96%)とシングルリーダーがCADを使った場合(97%)は、ほぼ同等の検出感度であるという報告が出ています。オランダのナイメーゲン大学のニコ・カーシメージャ²⁶⁾の報告では、シングルリーダー(39.4%)、ダブルリーディング(49.9%)、そしてCADを用いた場合(46.4%)を比較すると、ダブルリーディングのほうがまだ高く、CADはまだ足りないという報告です。その後、イタリアで、別のグループ²⁷⁾が、70の悪性の微小石灰化について読影実験がされています。この場合には、シングルリーダー(78.1%)とCADの効果(90.5%)を調べています。CADを用いることが明らかに効果があるという報告です。イギリスのロンドンのグループの研究²⁸⁾は、2005年にRadiologyに報告されています。シングルリーダー(90.2%)、CADを用いた場合(91.5%)、ダブルリーディング(98.4%)の検出感度を比較すると、これもダブルリーディング

のほうが特性がよいという報告です。しかし、最近の2006年になって、スコットランドのアーバディン大学のフィオーナ・ギルバード²⁹⁾の研究が報告されています。この研究では、非常に多くのケース(10,267例)と230件の乳癌のケースを用いて、シングルリーダーCAD(49.1%)とダブルリーディング(42.6%)の比較をしています。この結果によれば、ダブルリーディングよりも、CADを用いたシングルリーダーのほうが乳癌検出感度が高いと報告されています。このことから、ギルバードたちは、大がかりなCADの実験を行うべきである、クリニカルトライアルを行うべきであると結論しています。

最後は、臨床のプロスペクティブ研究です。最初の研究は、テキサスのフリーア³⁰⁾というプライベートプラクティスの医師が行ったのですが、2001年にRadiologyに報告されています。ここでは、1万人以上の受診者について、CADを利用することによって、乳癌の検出率が19.5%向上するが、再診率もほぼ同等の増加分があるということが報告されています。その後、10万人以上の受診者について、ピッツバーグ大学のデイビッド・ガー³¹⁾の報告が、Journal of the National Cancer Instituteに報告されています。これは約5万人ほどの受診者を、CADを用いる前に検出した結果と、CADを使った後の比較をしたのですが、もちろん患者が違いますから問題がありますが、検出率はほんのわずかだけ増加している(1.7%)という報告があります。ところが、UCLAのステーブ・フェイグ³²⁾は、このデータをさらに詳しく分析すると、ローボリュームの読影医師の扱った約3万人ほどの患者については、CADを用いることによって、乳癌の検出率は19.7%改善されていると報告しています。さらに、スタンフォードのバードウェル³³⁾は、CADを用いることによって、乳癌の検出率が7.4%改善されたと報告しています。サウスカロライナのプライベートプラクティスのカップルズ³⁴⁾の報告では、CADによって乳癌の検出率は16.1%増加しています。これは3年間の研究ですが、最初の1年間はCAD

なしの条件で、最後の2年間はCADを用いた結果を分析したものです。この報告の中で重要なのは、1 cm以下のインベーシブキャンサーの場合、検出率が164%向上しています。しかも乳癌が検出されたときの患者の年齢が53歳若くなっていると報告されています。さらに、メイヨークリニック³⁵⁾や、カリフォルニアのプライベートプラクティスのグループ³⁶⁾からも、CADを用いることによって、乳癌の検出率はそれぞれ7.6%および10.8%向上したと報告されています。

3. マモグラムにおける鑑別診断

マモグラムに含まれる異常陰影が検出されると、次は、果たして良性か悪性かということになります。微小石灰化については、これを区別するためのCADの開発が進んでいます。基本的なアルゴリズムとしては、検出された微小石灰化に対して、それらの特徴量を分析し、ニューラルネットワークを用いて悪性度の確率(%)を決めるという手法です。微小石灰化の特徴量としては、クラスターの面積と円形度、さらに微小石灰化の数、平均面積、平均体積、それらの標準偏差、そして形状に関する尺度^{37,38)}などです。これらの特徴量のうち、クラスターの面積と形状についての関係をプロットすると、良性と悪性の微小石灰化のデータはかなりオーバーラップしていますが、乳癌のほうが良性のものよりもクラスターの面積が比較的大きいことがわかります。このように、多くの特徴量はかなりオーバーラップしていますが、ニューラルネットワークを使うことによって、良性と悪性の区別がある程度できるようになります。

次に、ニューラルネットワークから得られる悪性度の確率が医師にとって役に立つかどうかを調べるために、オブザーバーの実験³⁸⁾が行われています。この実験では、病巣の位置を示すマモグラムと拡大撮影像を医師に提示します。ROC曲線を求めるために、医師は病巣について悪性度の程度を主観的に判断し、その確信度をマークします。この確信度は不連続な数値でも良いのです

が、最近ではコンピュータのモニターを利用して、アナログの連続値を用いることが多くなっています。医師ははじめにコンピュータの結果なしで、次にコンピュータの結果、悪性度(例えば80%など)を示す数字を見て、良性、悪性の判断をし、確信度を決定します。得られたROC曲線では、コンピュータの結果を用いないときの放射線科医の特性が一番低く、45度の直線に近い曲線が得られています。ROC曲線の面積(Az値)は、0.61です。そこで、医師にとっては、良性、悪性の区別がきわめて難しいことを示しています。一方、コンピュータの結果は、かなり高い特性(Az値0.80)が得られています。放射線科医は、コンピュータの結果を見ることによって、かなり改良(Az値0.75)できますが、まだコンピュータの結果よりも若干低いことを示しています。このことは、将来、医師がコンピュータの結果をより良く利用できるような方法を考案する必要があることを示していると思います。その1つの方法が、後述の類似画像です。

腫瘍の鑑別診断³⁹⁾については、腫瘍の形状、エッジグラジエントなどを調べて、良性か悪性かを区別します。悪性腫瘍の場合には、エッジグラジエントはいろいろな方向を向いていますので、ヒストグラムの幅が広がります。一方、良性腫瘍はヒストグラムの幅が狭くなりますので、ヒストグラムの幅に関する特徴量を用いることによって、良性、悪性の区別をすることが可能になります。しかし、良性、悪性腫瘍の特徴量もかなりオーバーラップしていますが、若干の相異もありますので、これらの特徴量を利用して、乳癌と良性腫瘍をある程度区別できる可能性を示しています。腫瘍の鑑別診断についても、コンピュータの決定した悪性度の確率に関するオブザーバーの実験が行われています。この結果によると⁴⁰⁾、コンピュータの特性(Az値0.90)が一番低いのですが、これを利用することによって、医師の特性(Az値)は、0.93から0.96まで改善されています。そこで、腫瘍陰影の良悪性の鑑別に関しては、コンピュータの特性と医師の相乗効果が起きている

ことを示しています。

4. マモグラムのデータベース

CAD に関する研究をするためには、できるだけ多くのケースを用いることが必要です。現在、データベースが幾つかありますが、その中で一番大きなデータベースは、南フロリダ大学⁴¹⁾の乳癌検診のためのデジタルデータベース(DDSM)です。この中には、914の乳癌、855の良性病巣、さらに695の正常例が含まれています。これらの中には、1箇所だけのマス、複数のマス、さらにマスと微小石灰化の混ざったもの、一箇所だけの微小石灰化、複数の微小石灰化など多くの例がありますので、CADの開発研究に用いることが可能になっています。

このデータベースの性質を見ると、腫瘍の寸法は、小さいものは5 mm 以下から、大きいものは50mm 以上まで分布しています。良性腫瘍の寸法と悪性腫瘍の寸法は、かなりオーバーラップしています。微小石灰化クラスターの寸法の分布を調べると、悪性のものの寸法よりも、良性のものの寸法は小さいものが多い傾向があります。また、微小石灰化の数の分布については、微小石灰化の数の少ないものは良性である可能性が強いことを示しています。腫瘍の寸法とコントラストの関係については、良性と乳癌はほとんどオーバーラップしています。一方、腫瘍の寸法と腫瘍の形状に関する不規則度との関係では、乳癌のほうが不規則度が大きい傾向を示しています。しかし、全体にはかなりオーバーラップしています。さらに、微小石灰化のクラスターの寸法と石灰化の数の関係や、クラスターの寸法とクラスターの円形度との関係は、いずれも幅広く分布していますが、かなりオーバーラップしています。

5. 鑑別診断における類似画像の役割と重要性

第4図は、マモグラムの腫瘍陰影の良悪性の鑑別について、類似画像が役に立つ可能性を示す例です。中央にある腫瘍陰影は未知の新しいケースとします。そこで、多くの良性、悪性の例を含む

大きなデータベースの中から未知のケースに似ている腫瘍陰影を探します。この例では、左右にそれぞれ2例ずつの良性腫瘍と乳癌が表示されています。これらと比較すると、ほとんどの方が賛成すると思いますが、この未知のケースは乳癌の病巣により似ていると思います。実際、これは乳癌の例です。したがって、このような類似画像の例を比較観察することによって、医師はこの未知のケースは乳癌らしいと推定できます。では、一体なぜ類似画像を利用するのかという疑問があります。この理由の1つは、医師は多数の臨床画像を見ることによって教育されているからです。多くの画像を見て、病気や正常例と関係するそれぞれに共通するパターンを覚えることが画像読影の教育の一部になっています。さらに、実際の臨床においては、ときおり未知の臨床例と既知の病巣を含む画像とを比較したくることがあります。そのためには、教科書や過去の臨床例などを見ることがあるといわれています。類似画像という概念は、人間が一般的に持っているものだと思います。しかし、医用画像の分野における類似画像は、今までに定量的に分析したり、評価されたことがありません。そこで、類似画像の“類似度”は、果たして主観的、客観的にどう定義できるのか、どうやって測れるのか、類似画像はどうやって見つけるのか、類似画像はどのくらい役に立つのかというような疑問が次々に問題点として出てきます。したがって、類似画像は重要な研究対象の1つと考えられます。そこで、シカゴの私どもの研究室では、最近、類似画像に関する研究⁴²⁻⁴⁵⁾を始めています。その例を、以下に述べます。

まずはじめに、放射線科医による主観的類似度の測定です。主観的類似度の決定の実験には、例えば第5図のように、2つの腫瘍陰影を1つのペアとして、3つのペアを用意します。それらについて放射線科医は、0 1.0のスケールによる連続確信度を用いて、類似の程度を判断します。スケールの左端(0)は“全く似ていない”、右端(1.0)は“殆ど同じに見えるほど似ている”に対応します。このスケールの適当なところをクリックする

第4図. 乳癌の検診における鑑別診断での類似画像の有用性を示す例です．中央の腫瘍陰影は未知の新しいケースとすると，大きなデータベースから，未知のケースに類似する良性腫瘍と乳癌の例を提示します．未知のケースは，乳癌の例に良く似ているので，医師は自信を持って乳癌と判断できます．

第5図. 腫瘍陰影のペアについて，放射線科医は0-1.0のスケールによる連続確信度を用いて，主観的類似度を決定することができます。スケールの左端(0)は“全く似ていない”，右端(1.0)は“殆ど同じに見えるほど似ている”に対応します．3つのペアについては，放射線科医による主観的類似度の平均値と標準偏差を示します．

と，主観的類似度が決定されます。このようにして，多くの医師によって得られた実験結果の平均

値は，一番上の腫瘍陰影ペアは 0.85 ± 0.03 ，下のペアになると， 0.56 ， 0.27 と類似度が下がってき

第6図. 客観的類似度の決定にはニューラルネットワークが用いられます。ニューラルネットワークは、主観的類似度と客観的特徴量との関係を学ぶために利用します。入力データは2つの陰影の特徴量で、教師信号は主観的類似度です。トレーニングのできたニューラルネットワークに未知のペアの特徴量を入力すると、出力として心理物理的類似度の値が得られます。

ます。この実験では、ペアの陰影の類似の程度の判断基準は、放射線診断の目的に対してであることを明確に医師に指示する必要があります。この判断基準はきわめて重要です。それは、この判断基準によって、主観的類似度が大きく変わる可能性があるからです。例えば、乳癌専門の放射線科医と一般の放射線科医では、類似度の判断が異なる場合があります。この理由は、専門医には良悪性の区別に利用できる淡い重要な陰影の特徴であっても、一般の医師には経験のない陰影の特徴であるために、その重要性に気がつかない場合があるからです。

このように主観的な類似度が決められたとすると、次は、物理的にはどうやって類似度を決めたらよいか、客観的に決めることができるかどうかです。これについては、第6図のようなニューラルネットワークを用いることが試みられています。ニューラルネットワークは、主観的類似度と客観的特徴量の関係を学ぶために用います。入力データとしては、与えられた2つの腫瘍陰影から、平均寸法、ピクセル値の標準偏差、円形度などの特徴量を定量化して、これを入力データとします。一方、教師信号としては、放射線科医から得られた主観的類似度の平均値を与えます。この

ようなデータを繰り返し用いることによって、ニューラルネットワークは、多くのペアの物理的特徴量と主観的類似度との関係を学びます。一度学ばると、未知のペアの画像の特徴量を入力すると、出力として得られるのが、心理物理的類似度の尺度ということになります。このようにして得られる心理物理的類似度の尺度は、どの程度の信頼性があるかを検討する必要があります。そこで、縦軸と横軸に、それぞれ心理物理的類似度の尺度と、放射線科医による主観的類似度の平均値をプロットします。その結果、主観的類似度と心理物理的類似度との相関値は 0.818 となっています。これは、かなり高い相関値と考えられます。そこで、ニューラルネットワークを用いて得られる心理物理的類似度の値^{42, 43)}は、放射線科医の主観的類似度と比較的良好な対応をしていることがわかります。

次に、主観的類似度の評価について、最近、別の手法⁴⁵⁾を検討しています。人間は、絶対値のスケールでどのぐらい似ているかという主観的判断ができる一方、2つのものを比較すると、より信頼性が高く、再現性の良い判断のされることが知られています。この手法は、一対比較法やツー・アルタナティブ・フォースド・チョイス(2 AFC)

第7図. 腫瘍のペアの相対的な類似度は、2つのペアを比べる一対比較法によって決定することができます。上のペアと下のペアを10人の放射線科医が2回ランダムに比べると、20回のうち16回は上のペアの方がより似ていると判断されます。より似ていると判断された回数によって、多くのペアの主観的類似度を相対的に比較できます。

法と言われています。この実験では、**第7図**のように、腫瘍のペアを2セット用意します。観察者は、上側の腫瘍のペアの似ている程度と下側の腫瘍ペアの似ている程度とを比較して、より似ているペアを選択します。この例の場合には、上のペアのほうがより似ていると判断する方が多いと思います。実際に、10人の放射線科医が2回ずつランダムに判断した結果、20回のうち16回は、上のペアのほうがより似ていると判断しています。そこで、より似ていると判断した回数によって、どちらのほうが類似度が高いかをランクづけることができ、多くのペアの主観的類似度を相対的に比較することが可能になります。

第8図は、8ペアの腫瘍陰影を用いて、すべての組み合わせの一対比較の実験を行った結果を示しています。横軸が絶対値による主観的類似度の値、縦軸が多くのペアを比べることによって得られた類似度ランキングで、これは相対値です。この結果は、非常に高い相関値(0.94)が得られてい

第8図. 腫瘍陰影のペアについての主観的類似度(絶対値)と類似度ランキング(相対値)との関係は高い相関値を示します。このことは、主観的類似度を評価する2つの測定法には一貫性のあること示しています。

ますから、絶対値で判断した主観的類似度と相対的に比較した類似度のランキングとは非常に対応が良く、一致していることを示しています。一方、微小石灰化のクラスターについても同様の一対比較の実験が行われています。その結果、主観的類似度とランキングの相対値の相関値は0.98という高い値が得られています。そこで、主観的類似度を評価する2つの測定方法には、一貫性のあることが示されています。

通常は、腫瘍のペアと微小石灰化のペアを比較することはありません。しかし、類似度の概念が一般性のある概念であるならば、それらの別々のペアがあったときに、どちらのほうがより似ているかを考えることは可能です。そこで、**第9図**のように、腫瘍のペアと微小石灰化のペアで、どちらがよく似ているかを比較⁴⁵⁾してみました。この例では、20回のうち16回は上のペアのほうが、より似ているという判断がされています。石灰化の4ペアと腫瘍の4ペアのすべての組み合わせを用いてランキングした結果も、主観的類似度のデータとかなり高い相関値(0.96)が得られています。このことから、人間が判断する医用画像の病巣に

第9図. 微小石灰化クラスターのペアと腫瘍のペアとの類似度は、一対比較法によって比べることができます。この例では、20回のうち16回は上のペアの方がより似ていると判断されています。

関する類似度を定量化することは可能であり、また、かなり一般性の強い概念と思われます。

おわりに

マモグラムによる乳癌検出のコンピュータ支援診断はようやく最近実用化が始まり、乳癌検診における画像診断に大きなインパクトを与えることが期待されます。しかし、現在のCADに関してはさらに改善する余地がありますし、重要と思われる問題点もあります。臨床的な研究に関しては、多くの異なった臨床の状況に応用すること、さらにCADの臨床の有用性をまったく疑いのないほど明確に示す証明が必要です。コンピュータによる検出特性については、大きなデータベースを利用して、さらに大きく改良することが要求されます。いままでの臨床実用のCADは検出を目的としていますが、今後は良性、悪性を区別する鑑別診断を導入することが必要です。一方、5年後にはコンピュータ支援診断は大きく成長している可能性^{46, 47)}があります。多分、各種のモダリティによる多くの異なった病巣に対するCADシ

ステムが開発され、検出特性は改善され、実用化が進展すると思います。また、PACSは現在多くの病院での実用が始まっていますが、今後はさらに成長して、PACSの一部にCADのソフトウェアが導入されると思われます。そこで、CADは、PACSに含まれる巨大なデータベースからの類似画像を利用することによって、診断支援だけでなく、医師の生涯教育用のツールとしても使われるようになると思います。最終的には、CADは医療報酬を含む標準的な診断業務になると予測されます。

【文 献】

- 1) Byers T, Barerra E, Fontham ETH, et al: A midpoint assessment of the American cancer society challenge goal to halve the U.S. cancer mortality rates between the years 1990 and 2015. *Cancer*, 107: 396 405, 2006
- 2) Doi K, Giger ML, MacMahon H, et al: Computer-aided diagnosis: Development of automated schemes for quantitative analysis of radiographic images. *Seminars in Ultrasound, CT and MRI*, 13: 140 152, 1992
- 3) Doi K, MacMahon H, Giger ML, Hoffmann KR (editors): *Computer Aided Diagnosis in Medical Imaging* (Elsevier, Amsterdam), pgs 3 560, 1999
- 4) Doi K, MacMahon H, Katsuragawa S, Nishikawa RM, Jiang Y: Computer-aided diagnosis in radiology: potential and pitfalls. *European Journal of Radiology*, 31: 97 109, 1999
- 5) Doi K: Computer-aided diagnosis in digital chest radiography. In: *Advances in Digital Radiography: RSNA Categorical Course in Diagnostic Radiology Physics 2003 Syllabus*, 2003, pp. 227 236
- 6) Doi K: Overview on research and development of computer-aided diagnostic schemes. *Seminars in Ultrasound, CT and MRI*, 25: 404 410, 2004
- 7) Doi K: Current status and future potential of computer-aided diagnosis in medical imaging. *British J. Radiology, Special Issue*, 78: S 3 S 19, 2005
- 8) Doi K: Diagnostic imaging over the last 50 years: Research and development in medical

- imaging science and technology. Physics in Medicine and Biology, 50th Anniversary Issue , 51 : R 5 R27 , 2006
- 9) 土井邦雄：コンピュータ支援診断(CAD)：基礎概念，現状，および将来の可能性．日獨医放，48：8 20，2003
 - 10) 土井邦雄：CAD 開発における国際的歴史と実用化の世界情勢．インナービジョン，11：1 4，2004
 - 11) Kundel H L, Nodine C F, Carmody D: Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Investigative Radiology, 13 : 175 181 , 1978
 - 12) Chan HP, Doi K, Vyborny CJ, Schmidt RA, Metz CE, Lam KL, Ogura T, Wu Y, MacMahon H: Improvement in radiologists' detection of clustered microcalcifications on mammograms: The potential of computer-aided diagnosis. Investigative Radiology, 25 : 1102 1110 , 1990
 - 13) Warren Burhenne LJ, Wood SA, D'Orsi CJ, et al: Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology , 215 : 554 562 , 2000
 - 14) Birdwell RL, Ikeda DM, O'Shaughnessy KF, Sickles EA: Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. Radiology , 219 : 192 202 , 2001
 - 15) Brem RF, Baum J, Lechner M, et al: Improvement in sensitivity of screening mammography with computer-aided detection: A multiinstitutional trial. AJR , 181 : 687 693 , 2003
 - 16) Evans WP, Warren Burheene LJ, Lauri L, O'Shaughnessy KF, Castellino RA: Invasive lobular carcinoma of the breast: Mammographic characteristics and computer-aided detection. Radiology , 225 : 182 189 , 2002
 - 17) Baker JA, Rosen EL, Lo JY, et al: Computer-aided detection(CAD) in screening mammography: Sensitivity of commercial CAD systems for detecting architectural distortion. AJR , 181 : 1083 1088 , 2003
 - 18) Malich A, Sauner D, Marx C, et al: Influence of breast lesion size and histologic findings on tumor detection rate of a computer-aided detection system. Radiology , 228 : 851 856 , 2003
 - 19) Destounis SV, DiNitto P, Logan-Yong W, et al: Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience. Radiology , 232 : 587 584 , 2004
 - 20) Ikeda DM, Birdwell RL, Shaughnessy KF, et al: Computer-aided detection output on 172 subtle findings on normal mammograms previously obtained in women with breast cancer detected at follow-up screening mammography. Radiology , 230 : 811 819 , 2004
 - 21) Butler SA, Gabbay RJ, Kass DA, et al: Computer-aided detection in diagnostic mammography: Detection of clinically unsuspected cancers. AJR, 183 : 1511 1515 , 2004
 - 22) Soo MS, Rosen EL, Xia JO, et al: Computer-aided detection of amorphous calcifications. AJR, 184 : 887 892 , 2005
 - 23) Brem RF, Rapelyea JA, Zisman G, et al: Evaluation of breast cancer with a computer-aided detection system by mammographic appearance and histopathology. Cancer , 104 : 931 935 , 2005
 - 24) Brem RF, Hoffmeister JW, Rapelyea JA, et al: Impact of breast density on computer-aided detection for breast cancer. AJR, 184 : 439 444 , 2005
 - 25) Ciatto S, Del Turco MR, Risso G, et al: Comparison of standard reading and computer-aided detection(CAD) on a national proficiency test of screening mammography. European Journal of Radiology , 45 : 135 138 , 2003
 - 26) Karssemeijer N, Otten JDM, Verbeek ALM, et al: Computer-aided detection versus independent double reading of masses on mammograms. Radiology , 227 : 192 200 , 2003
 - 27) Lauria A, Palmiero R, Forni G, et al: A study on two different CAD systems for mammography as an aid to radiological diagnosis in the search of microcalcification clusters. European Journal of Radiology , 55 : 264 269 , 2005
 - 28) Khoo LAL, Taylor, P, Given-Wilson RM: Computer-aided detection in the United Kingdom national breast screening programme: Prospective study. Radiology , 237 : 444 449 , 2005
 - 29) Gilbert FJ, Astley SM, McGee MA, et al: Single reading with computer-aided detection and

- double reading of screening mammograms in the United Kingdom national breast screening program. *Radiology*, 241 : 47 53, 2006
- 30) Freer TW, Ulissey MJ: Screening mammography with computer-aided detection: Prospective study of 12, 860 patients in a community breast cancer. *Radiology*, 220 : 781 786, 2001
- 31) Gur D, Sumkin JH, Rockette HE, et al: Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *Journal of the National Cancer Institute*, 96 : 185 190, 2004
- 32) Feig SA, Sickles EA, et al: Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *Journal of the National Cancer Institute*, 96 : 1260 1261, 2004
- 33) Birdwell RL, Bandodkar P, Ikeda DM: Computer-aided detection with screening mammography in a university hospital setting. *Radiology*, 236 : 451 457, 2005
- 34) Cupples TE, Cunningham JE, Reynolds JC: Impact of computer-aided detection in a regional screening mammography program. *AJR*, 185 : 944 950, 2005
- 35) Morton MJ, Whaley DH, Brandt KR, et al: Screening mammograms: Interpretation with computer-aided detection-prospective evaluation. *Radiology*, 239 : 375 383, 2006
- 36) Dean JC, Ilvento CC: Improved cancer detection using computer-aided detection with diagnostic and screening mammography: Prospective study of 104 cancers. *AJR*, 187 : 20 28, 2006
- 37) Jiang Y, Nishikawa RM, Wolverton DE, Metz CE, Giger ML, Schmidt RA, Vyborny CJ, Doi K: Malignant and benign clustered microcalcifications: Automated feature analysis and classification. *Radiology*, 198 : 671 678, 1996
- 38) Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Giger ML, Doi K: Improving breast cancer diagnosis with computer-aided diagnosis. *Academic Radiology*, 6 : 22 33, 1999
- 39) Huo Z, Giger ML, Vyborny CJ, et al: Computerized classification of benign and malignant masses on digitized mammograms: a robustness study. *Academic Radiology*, 7 : 1077 1084, 2000
- 40) Huo Z, Giger ML, Vyborny CJ, Metz CE: Breast cancer: effectiveness of computer-aided diagnosis-observer study with independent database of mammograms. *Radiology*, 224 : 560 568, 2002
- 41) Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer P: Current status of the digital database for screening mammography. In: *Digital Mammography*, Kluwer Academic Publishers, Dordrecht, 1998, pp 457 460
- 42) Li Q, Li F, Shiraishi J, Katsuragawa S, Sone S, Doi K: Investigation of new psychophysical measures for evaluation of similar images on thoracic CT for distinction between benign and malignant nodules. *Medical Physics*, 30 : 2584 2593, 2003
- 43) Muramatsu C, Li Q, Suzuki K, Schmidt RA, Shiraishi J, Newstead GM, Doi K: Investigation of psychophysical measure for evaluation of similar images for mammographic masses: Preliminary results. *Medical Physics*, 32 : 2295 2304, 2005
- 44) Muramatsu C, Li Q, Schmidt RA, Suzuki K, Shiraishi J, Newstead GM, Doi K: Experimental determination of subjective similarity for pairs of clustered microcalcifications on mammograms: Observer study results. *Medical Physics*, 33 : 3460 3468, 2006
- 45) Muramatsu C, Li Q, Schmidt RA, Shiraishi J, Suzuki K, Newstead GM, Doi K: Determination of subjective similarity for pairs of masses and pairs of clustered microcalcifications on mammograms: Comparison of similarity ranking scores and absolute similarity ratings. *Medical Physics*, in press, 2007
- 46) Doi K: Computer-aided diagnosis moves from breast to other systems. *Diagnostic Imaging*, May, pp 37 40, 2007
- 47) Doi K: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Journal of Computerized Medical Imaging and Graphics*, 31 : 198 211, 2007