5. 系列依存性を利用したニューラルネットワークによる乳房腫瘍の良悪性の鑑別

蔡 篤 儀・藤 田 広 志*・堀 田 勝 平**

遠 藤 登喜子***・木 戸 長一郎**・石 垣 武 男***

岐阜工業高等専門学校

*岐阜大学工学部

**愛知県がんセンター病院

***名古屋大学医学部

1. ま え が き

食事や生活習慣の欧米化に伴い、日本における乳がんの増加は甚だしく、今世紀末には女性ががんの第一位になると予測されている。乳がんは、早期診断によって治療の期待できるがんの一つでもある。1987年から老健法による集団検診が行われてきた。しかし、この場合は膨大な読影という課題を解決する必要がある。

このような状況下で、X線のコンピュータ支援診断の研究が盛んになってきた¹⁻⁶⁾。例えば、I.I.-TV方式マンモ専用のDR装置と陰影を検出するマンモグラム診断支援ソフトウェアが開発されている⁹⁾。しかし、このソフトには、検出された陰影の良悪性を判別するアルゴリズムが含まれていない。

そこで我々は、ニューラルネットワークによるパターン認識の手法⁷⁻¹⁰⁾を、検出された腫瘍候補の良悪性の判別に適用することを試み、その有用性を検討した¹¹⁾。この方法では、画像データを直接ネットワークに入力し、平均75%の認識率を得た。

本稿では、認識率を向上させるため、内田らが提案した情報理論における系列依存性の手法¹²⁻¹⁴⁾を応用して、画像の特徴量を抽出し、ニューラルネットワークの入力データとする方法について紹介する¹⁵⁾。

2. 系列依存性

1列の記号系列がm種類の記号からなるとすれば、その最大エントロピー H_0 は、それら記号の発生確率がすべて等しく p である場合で、

$$H_0 = \sum p \log_2 \frac{1}{p} = m \left(\frac{1}{m} \log_2 m \right) = \log_2 m \quad (1)$$

で与えられる。この与えられた最大エントロピーを0次推定値と呼ぶ^{16,17)}。

1次推定値 H_1 は、その記号系列のエントロピーと定義し、それぞれの記号の発生確率を p_i とすると

$$H_1 = \sum p_i \log_2 \frac{1}{p_i} \quad (2)$$

で表される。

そして次に一対の連続する記号 (ダイグラム) を1つのまとめた記号であるとみなし、その発生確率を p (digram) とすると、その記号系列のエントロピーは、

$$H(\text{digram}) = \sum p(\text{digram}) \log_2 \frac{1}{p(\text{digram})} \quad (3)$$

となる。2次推定値 H_2 は、 $H(\text{digram})$ と H_1 との差をとれば求めることができる。すなわち、

$$H_2 = H(\text{digram}) - H_1. \quad (4)$$

3次の推定値 H_3 についても同様に、3つの連続記号を1つのまとめた記号 (trigram) とみなし、同様にその記号系列のエントロピー $H(\text{trigram})$ を計算し、 H_3 を

$$H_3 = H(\text{trigram}) - H(\text{digram}) \quad (5)$$

で求めることができる。一般化すると、次の式に帰納できる：

$$H_N = H(N\text{-gram}) - H(N-1\text{-gram}) \quad (6)$$

ここで、 $N\text{-gram}$ の発生確率を $p_i(N\text{-gram})$ とし、

$$H(N\text{-gram}) = \sum p_i(N\text{-gram}) \log_2 \frac{1}{p_i(N\text{-gram})} \quad (7)$$

と定義する。詳細については、ほかの文献を参照されたい^{17,18)}。

このようにして計算された画像の系列依存性を、一種の特徴量とし、ニューラルネットワークの入力データにする。

3. 実験方法

3.1 対象画像データ

本実験で使用したすべての画像データは、マンモ用 DR 装置を用いて、愛知県がんセンター病院で撮影されたデジタルマンモ画像（マトリックスサイズ：1024×1024，階調：10ビット＝1024）である。このオリジナル画像から腫瘍候補の周辺256×256画素を切り出し，“抽出画像”として解析に使用した。腫瘍候補数は，良性20例と悪性20例の計40例である。なお，腫瘍でないもの（病

Fig. 1 Basic structure of our artificial neural network.

変でない部分）を腫瘍として誤認識するケースを想定し，そのようなサンプルも良性例に入れた。

3.2 ニューラルネットワークの構成

ニューラルネットワーク（階層型）には，パーソナルコンピュータ（PC-9801 DS2，NEC）を用いるシステム（Neuro-07）を使用した。学習は，バックプロパゲーション法によった。本システムの層構造は，入力層，中間層，出力層の3層構造である。学習演算に必要なパラメータとしては，実験によって次の値を使用した。すなわち，荷重・しきい値初期化パラメータ： $-0.5 \sim +0.5$ ，修正係数：0.25，忘却係数：1である。

本実験で用いたニューラルネットワークの基本構成を Fig. 1 に示す。入力層のユニット数は1000個にし，出力層は良性と悪性の2ユニットとした。中間層のユニット数は可変とし，最適値を実験で求めた。入力層のユニットには，後述する系列依存性の手法によって得られた画像特徴量を入力する。

3.3 系列依存性の手法による画像の特徴抽出

本研究で使用するパーソナル・ニューロコンピュータの入力層のユニット数には制約があるため（最大1024個まで），背景トレンド除去後の256×256サイズの画像から中心部の200×200の ROI（region of interest）を取り出し，量子化レベルを16とした。

系列依存性の手法による画像特徴の抽出の概念を Fig. 2 に示す。まず，画像の横方向の一番上の系列を取り出し，式(1)～式(7)を用いて $H_0 \sim H_4$ を計算する。次に，二番目の系列も同様な方法で計算する。このように次々

Fig. 2 Extraction of feature vectors using sequential-dependence method.

Fig. 3 Processing flow of mammograms. Preprocessing is the part within the dashed lines.

(a)

(b)

Fig. 4 Images of (a) benign tumor and (b) malignant tumor in mammograms.

とエントロピーを求め、最終的に、総数 $5 \times 200 = 1000$ 個の画像特徴量を抽出できる。

3.4 実験のフロー

実験は Fig. 3 のような処理フローで行った。ニューラルネットワークによる解析では、40例を良悪性同等にランダムに2等分し、一方を学習用データに、他方を認識用データとして使用した。学習回数は1000回で行った。

前処理の段階での陰影周辺のトレンド除去の手法を次に示す¹⁾。まず、 127×127 画素のマスクサイズの単純平滑化フィルタによる低周波画像（トレンド成分）を作成する。そして、この画像を抽出原画像から減算し、 256×256 サイズの原画像の平均値を加算する操作を行う。

4. 結果と考察

トレンド除去の処理後の典型的な良悪性腫瘍画像の例を Fig. 4 (a)と(b)に、またそれぞれの正規化した系列依存性分布を Fig. 5 (a)と(b)に示す。このように、画像の有効な特徴量を取り出すことができる。

Table 1 に中間層のユニット数を変化させたときの認識率を示す。30ユニットのときが最も認識率が高く85%である。他の場合は、すべて80%である。この結果から、中間層のあるユニット数で認識率の最大値が存在し、必ずしもユニット数が増加するほど認識率が向上しない現象がみられる。系列依存性の手法による認識結果は、前回の画像を直接入力する方法に比べて、約10%の認識率の向上ができた。ニューラルネットワークの荷重・しきい値初期化パラメータおよび修正係数の変化による効果をも調べたが、認識率は、ネットワークの諸パラメー

Fig. 5 Normalized sequential-dependence distributions (a) and (b) for the images shown in Fig. 4(a) and (b), respectively.

タによらず一定で、非常に安定していることがわかった。

この安定性は、新しい種類の学習あるいは認識データが取り込まれたとき威力を発揮し、いろいろな特性の異なる入力データが現れても、学習時のパラメータを変更することなく、すなわち、システムの特徴のうち人間が変えてやらなければならない点を毎回変更することなく、学習・認識が行えることになる。このように、認識率の向上とそのパラメータに対する安定性において、特徴抽出法として系列依存性を導入した意義があったといえる。

5. む す び

系列依存性を利用したニューラルネットワークによる乳房腫瘍の良悪性の鑑別法を説明し、その有効性を示した。しかしながら、実用化のために認識率をさらに向上させるには、十分な学習と認識試験を行うための画像デ

Table 1 Recognition rates for various numbers of hidden units in the hidden layer.

Number of Hidden Units	Recognition Rate (%)
20	80
30	85
50	80
60	80
75	80
100	80
110	80

ータの増加や前処理の改良などの検討が必要である。なお、本稿で述べたのとは異なった手法により、最近さらに認識率の向上を実現できている^{19,20)}。

参 考 文 献

- 1) Dhawan AP, Buelloni G and Gordon R: Enhancement of mammographic features by optimal adaptive neighborhood image processing, IEEE Trans. Med. Imaging, 5, 8-15 (1986).
- 2) Lai SM, Li X and Bischof WF: On techniques for detecting circumscribed masses in mammogram, IEEE Trans. Med. Imaging, 8, 377-386 (1989).
- 3) Brzakovic D, Luo XM and Brzakovic P: An approach to automated detection of tumors in mammograms, IEEE Trans. Med. Imaging, 9, 233-241 (1990).
- 4) Giger ML, Yin F-F, Doi K, et al: Investigation of methods for the computerized detection and analysis of mammographic masses, Proc. SPIE, 1233, 183-184 (1990).
- 5) Yin F-F, Giger ML, Doi K, et al: Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images, Med. Phys., 18, 955-963 (1991).
- 6) Endo T, Kido C, Horita K, et al: Clinical evaluation of assistant diagnostic system for mammograms using the auto-analyzing method, Radiation Medicine, 10, 50-54 (1992).
- 7) Rumelhart DE, Hinton GE and Williams RJ: Learning representation by backpropagation errors, Nature, 323, 533-536 (1986).
- 8) Udpa L and Udpa S: Neural networks for the classification of nondestructive evaluation sig-

〔第33回画像部会報告〕 5. 系列依存性を利用したニューラルネットワークによる乳房腫瘍の良悪性の鑑別 (蔡・他)

- nals, IEE Proc-F, 138, 41-54 (1991).
- 9) Floyd Jr CE: An artificial neural network for SPECT image reconstruction, IEEE Trans. Med. Imaging, 10, 485-487 (1991).
- 10) Qin SZ, Su HT and McAvog TJ: Comparison of four neural net learning methods for dynamic system identification, IEEE Trans. Neural Networks, 3, 122-130 (1992).
- 11) 藤田広志, 堀田勝平, 遠藤登喜子, 他: ニューラルネットワークによるマンモグラフィ腫瘍陰影の良悪性の判別, Med. Imag. Tech., 10, 126-129 (1992).
- 12) 内田 勝, 蔡 篤儀, 小島克之: 画像の系列依存性による評価, 医画情誌, 1, 16-21 (1984).
- 13) Uchida S and Tsai DY: Evaluation of radiographic images by using sequential dependency, Jpn. J. Appl. Phys., 25, 1252-1255 (1986).
- 14) Tsai DY, Kojima K and Uchida S: A quantitative analysis of the effects of image processing in computed radiography using sequential-dependence method, Proc. of 12th Triennial World Congress of IMEKO XII, 861-866 (1991).
- 15) Tsai DY, Fujita H, Horita K, et al: Breast tumor classification by neural network fed with sequential-dependence factors to the input layer, IEICE Trans. INF. & SYST., E76-D(8), 956-962, (1993).
- 16) Miller GA and Frick FC: Statistical behavioristics and sequences of response, Psycho. Rev., 56, 2029-2034 (1949).
- 17) 内田 勝: 放射線画像工学, オーム社, 東京 (1986).
- 18) 内田 勝: 放射線画像工学実験, オーム社, 東京 (1985).
- 19) 蔡 篤儀, 坂本武彦, 藤田広志, 他: 統計的特徴を利用したニューラルネットワークによる乳房腫瘍の判別, 1993年電子情報通信学会春季大会予稿集, 6-192, (1993).
- 20) Tsai DY, Fujita H, Horita K, et al: Discrimination of breast tumors in mammograms by means of artificial intelligence: a revised approach, (submitted to Phys. Med. Biol.)