

# 3. 大量の臨床画像データの活用における問題

## — 4万例を超える臨床画像データの処理を行った CADシステムの開発・運用経験から

増谷 佳孝\*<sup>1</sup>/野村 行弘\*<sup>2</sup>/三木総一郎\*<sup>3</sup>/根本 充貴\*<sup>2</sup>  
花岡 昇平\*<sup>2</sup>/吉川 健啓\*<sup>3</sup>/林 直人\*<sup>3</sup>

\*1 広島市立大学情報科学研究科 \*2 東京大学医学部附属病院放射線科 \*3 東京大学医学部附属病院22世紀医療センター

### CAD技術の研究開発 における問題

近年、大量データに基づいた新しい知見の獲得や高性能な知的システムの構築に関する事例が、科学技術のみならずビジネスなどのさまざまな分野において報告されている。大量データの例としては、インターネット上の文書データ、eコマースや交通機関の使用履歴、気象や天文などのさまざまな科学技術データ、機密性の高い医療記録まで実に多岐にわたる。これらの大量データは「ビッグデータ」と呼ばれ、近年の科学技術の発展におけるキーワードとなっている<sup>1)</sup>。

医用画像の分野においては、アルツハイマー病の疾患解析のためのMRIデータベースであるADNI<sup>2)</sup>など、多くの画像データベース構築の試みがなされてきた。その中で、コンピュータ支援診断(computer-assisted detection/diagnosis:CAD)技術の研究開発者にとっては、大量の画像データをいかに入手して利用するかが、今なお重要な課題の一つである。CAD技術の基盤となるパターン認識・機械学習の観点からは、サンプルデータは多いほど良く、その量が性能や信頼性を決定づける。しかしながら、臨床施設以外の研究開発者にとって大量の医用画像データを得る手段は、ごく限定されている。さらに、データを入手できたとしても、研究レベルでは成功するもの

実用化には至らない事例が数多く見られ、すでに市場に投入されている商用システムについても、その効果は限定的であるように見える。その一方で、膨大な量の臨床画像が臨床施設のモダリティによって日々生成され、診療目的で使用された後、PACSに蓄積され続けている。臨床施設に所属するCAD研究開発者にとっても、これらの潜在的価値のすべてを引き出すのはそれほど容易ではない。

以上のような現状を踏まえ、臨床画像データを用いたCAD技術の研究開発においては、以下の3つの問題に留意する必要があると考えられる<sup>3)</sup>。

#### 1. 学習用データベース構築と更新

CAD技術は、基本的にパターン認識技術における「教師あり学習」に基づいており、データベースの構築には、単に画像データの収集のみならず、教師信号としての病変情報を入力する必要がある。しかし、これらはデータが大量になるほど高コストとなるため、そのほかの臨床情報の入力と同時に効率的なインターフェイスが必要となる。また、MRIの場合、CTに比べ、モダリティメーカー・機種間・撮像法によるデータの性質の差異は大きい。加えて撮像技術の進歩が著しいため、装置の更新時にも性質の変化がある。すなわち、一時期に大量データを収集して使い続けるのではなく、データの変化に応じた逐次更新が必要である。

#### 2. アルゴリズム・ソフトウェアの研究開発環境

学習処理を含むアルゴリズム上の試行錯誤を、最新データで効率的に行う環境が必要である。特に、大量のデータを使用して学習処理を行う場合、大容量のストレージや高性能計算を行う機器が必須となる。具体的には、学習処理におけるパラメータのチューニングのための最適化処理において、バッチ処理や並列化によりデータベースに効率的にアクセスし、処理を行うことが重要である。

#### 3. 臨床現場での使用における性能の維持・改善

臨床現場で使用するCADの結果表示のために、既存の読影環境と共存して利用できるようにインターフェイスが重要である。また、ある時点で収集したデータベースに基づいて開発したソフトウェアを使用し続けると、1. で述べたような機種の変更・更新によるデータの性質変化が性能を低下させる可能性がある。これを避けるためには、日々の臨床画像データの処理結果とその評価を研究開発に利用し、性能の改善・維持のためのモニタリングや調整を継続的に行うことが必要である。

以上の観点から、われわれは臨床施設内において、CADの技術およびシステムの研究開発と実運用を効率的に行う環境を検討し、設計・開発と実運用