

## 総括研究報告書

1. 研究開発課題名：医用知能情報システム基盤の研究開発
2. 研究開発代表者： 大江和彦（国立大学法人東京大学大学院医学系研究科・医学部附属病院）
3. 研究開発の成果

本研究の全体計画における目標は「医学研究と診療支援のための医用人工知能（AIM）基盤サービスの構築」である。具体的には（A）医用人工知能研究 DB（AIM-DB）の構築とそれによる類似症例高速検索機構の開発、（B）臨床データと医療知識のマッピング、（C）時系列データマイニングによる臨床 DB からの知識獲得、（D）機械学習 DL（ディープラーニング）による医療画像からの自動所見抽出、を開発すること、またその情報基盤を社会に提供することである。

初年度である 2015 年度は、（A）AIM-DB の構築として、東大、島根大、秋田大、浜松医大の 4 大学病院の SS-MIX2 標準ストレージからデータを匿名化して収集する匿名化システムを開発し、匿名化後 SS-MIX2 ストレージと AIM-DB を生成できるシステムを構築した。これにもとづき 4 大学病院での SS-MIX2 標準ストレージから過去 5 年分のデータを SS-MIX2 標準ストレージからデータ収集、匿名化処理後に AIM-DB に蓄積しつつある。また前記システムの匿名化後 SS-MIX2 ストレージから医療機関別統計情報収集システムを開発し、これを可視化するソフトのプロトタイプを開発し、4 大学病院のデータに対して並列動作できることを確認した。また 4 大学病院から本事業研究目的で VPN アクセスできる環境を構築した。

次に類似症例高速検索機構の開発として、東大病院の SS-MIX2 データを合田らが超高速データベースエンジンを用いてデータベース化し、複数の患者属性による検索を実現する Web ベースの検索インタフェイスと検索結果を時系列情報として可視化するプロトタイプを開発した。Web ベースのインタフェイス上に於いて、患者基本属性条件として性別ならびに年齢を、臨床条件として病名、検査値、処方薬の組み合わせを過去事象、注目事象、その後の事象の 3 つの事例系列区分に分けてそれぞれ指定することにより、匿名化 SS-MIX2 データから生成したデータベースに対して、検索処理を実行し、その結果得られた時系列情報を可視化することに成功した。

（B）臨床データと医療知識のマッピングに関しては、対象疾患として（1）循環器疾患領域、（2）糖尿病・代謝内科領域、（3）腎臓内科領域、（4）神経内科領域の 4 領域を選定し、臨床医学オントロジーに出現する 642 個（のべ 4,718 個）の異なる異常状態概念のカテゴリを「異常状態がどのように観測されるか」の視点により 5 個のカテゴリに分類した。

（C）時系列データマイニングによる臨床 DB からの知識獲得の準備として、過去 5 年間での検査オーダー実施数を病院情報システムから抽出し、実施数が多い、ホジキン病、慢性 C 型肝炎、慢性関節リウマチ、SLE をその対象と選定した。これらの 4 疾患について、SS-MIX のデータサーバーから検査データ、処方、記載についての抽出を完了し、①時系列頻出パターンマイニング法、②時系列予測法の 2 つのベースライン手法について実装を行った。

（D）医療画像からの自動所見抽出に関しては、ディープラーニングを高速に実施する GPU 計算機を導入し、200 枚程度での公開された胸部画像データでのパラメータ調整など事前処理準備を行った。また東大病院で倫理審査承認され、胸部 X 線画像データと所見データ（DPC 入院時傷病名情報）を匿名化して収集し、約 13,000 件の呼吸器内科の胸部 X 線正面画像データおよびうち 6,300 件の入院時傷病名にもとづく胸部診断情報を付与したデータセットを生成し、ディープラーニング処理系に取り込んだ。

以上。