

平成28年度厚生労働科学研究費補助金  
(政策科学総合研究事業(臨床研究等ICT基盤構築研究事業))  
総括研究報告書

「電子カルテ情報をセマンティクス(意味・内容)の標準化により分析可能な  
データに変換するための研究」

研究代表者	宮本 恵宏	国立循環器病研究センター循環器病統合センター・センター長
分担研究者	竹村 匡正	兵庫県立大学大学院応用情報科学研究科教授
	中村 文明	国立循環器病研究センター循環器病統合情報センター室長
	興梠 貴英	自治医科大学准教授
	中山 雅晴	東北大学病院メディカルITセンター教授
	的場 哲哉	九州大学病院講師
	小室 一成	東京大学大学院医学系研究科教授
	斎藤 能彦	奈良県立医科大学教授
	安田 聡	国立循環器病研究センター副院長・心臓血管内科部門長
	穴戸 稔聡	国立循環器病研究センター研究推進支援部部長
	西村 邦宏	国立循環器病研究センター循環器病統合情報センター室長
	平松 治彦	国立循環器病研究センター情報統括部部長
	上村 幸司	国立循環器病研究センター研究推進支援部研究情報基盤管理室室長
	辻田 賢一	熊本大学大学院生命科学研究部教授
	宇宿 功市郎	熊本大学医学部附属病院教授

研究要旨 日本循環器学会の事業である「臨床効果データベース」は、医療コストがかかる疾患・治療(心筋梗塞・狭心症とその病態に対するステント治療、重症心不全とそれに対する再同期療法(CRT))と循環器領域で特にその重要性が指摘されている疾患(急性心不全など)を抽出し、医療の質とその妥当性を検証するため時間軸を念頭においたデータベースである。本研究ではこのデータベースを用いて自然言語解析による診断判定システムの構築をおこなう。

#### A. 研究目的

高齢化社会の中にある我が国をはじめとする先進諸国において、循環器疾患が急増している。循環器疾患は再発を繰り返し徐々に進行していくという臨床経過をたどることが多い。例えば、虚血性心疾患では再発・入院を繰り返して終末像として心不全を呈することがしばしばある。そのため循環器疾患においては、Major Adverse Cardiac Event(MACE)とよばれる主要有害心血管イベントを発生させないための再発予防が大事である。循環器疾患の新規治療法の開発目標として、MACEの発生減少を目標としたものを開発することも考えられる

が、MACEを判断するためには担当した臨床医の判断が診療録を読み返し判断するしかない。そのため、レセプト/DPCなどの診療報酬請求情報を使用した分析、または電子カルテ情報を用いてビッグデータの分析においては、MACEなどのイベントをアウトカムにした研究をすることができないという限界がある。本研究では、電子カルテの記事情報から自然言語処理を活用して自動的にMACEであると判断するためのシステムを開発し、電子カルテ情報を用いたMACEのビッグデータ分析を行うためのシステムを開発する。日本語で記述される電子カルテからの臨床データベースにおいては初めての試み

である。

日本循環器学会の事業で実施している医療コストがかかる疾患・治療（心筋梗塞・狭心症とその病態に対するステント治療、重症心不全とそれに対する再同期療法（CRT））と循環器領域で特にその重要性が指摘されている疾患（急性心不全など）を抽出し、医療の質とその妥当性を検証するため時間軸を念頭においたデータベースである「臨床効果データベース」を用いて自然言語解析による診断判定システムの構築をおこなうことを目的としている。

## B. 研究方法

疾患コホート研究であり、虚血性心疾患、急性心不全の患者を対象とする。対象施設は、国立循環器病研究センター、東京大学、自治医科大学、自治医科大学さいたま医療センター、東北大学、九州大学にてデータの収集を行う。臨床効果データベースから、患者基本情報、診断名、入退院情報、経時的な内服薬、経時的な臨床検査情報、経時的な生理検査情報、経時的な心臓カテーテル検査情報を取得する。別途、電子カルテの記事情報を、臨床効果データベースと同じ匿名化番号にて匿名化した ID にて連結可能匿名化して受け取り、臨床効果データベースのデータと連結を行うことにより、電子カルテの記事情報と臨床データの結合を行う。さらに、入退院情報、検査結果等のデータから MACE による入院かどうかの判断を行う。これらの MACE のそれぞれのイベントに対して、ICD10 を用いて病名のコーディングを行う。次に、電子カルテの記事の医療用語を傷病名、愁訴、身体所見、検査、治療のそれぞれに分類を行う。これらの医療用語の出現と MACE の関係を機械学習（サポートベクターマシンやディープラーニング等）の手法を用いて学習を行い、電子カルテ記事から MACE かどうかを判定する予測モデルを作成する。

倫理面への配慮として、協力病院からデータを収集する際に、連結可能匿名化とし、個人識別情報および対応表は施設管理者の保管元、施設外に持ち出さないように厳格に管理する。

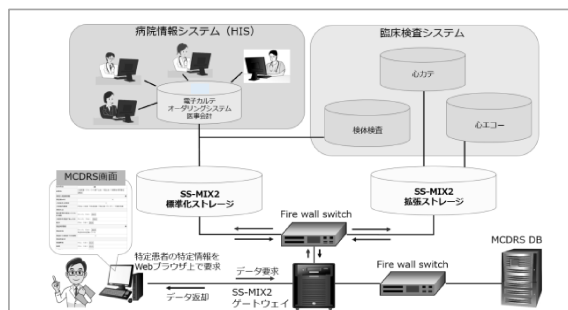
## C. 研究結果

本年度は、SS-MIX2 拡張ストレージに心電図、心臓超音波、心臓カテーテル検査結果の標準項目と形式とを定めた日本循環器学会標準出

力フォーマット（Standard Export data forMAT:SEAMAT）を用いて、国立循環器病センター、東北大学、自治医大、自治医大付属さいたま医療センター、東京大学、九州大学のデータを実際に格納する準備を行った。今回過去データも対象とするため CSV ファイルからデータ転送するプログラムを開発した。心電図については心電図サーバから SEAMAT 形式で SS-MIX2 拡張ストレージに書き出す準備が整った。（図 1）

## 多目的臨床データレポジトリシステム

図 1 臨床効果データベース



臨床効果データベースの情報を利用して人工知能に学習をさせる。

（MCDRS）（図 2）を用いて症例登録画面の作成を開始した。

図 2 多目的臨床データレポジトリシステム  
症例登録の画面例（各医療機関の登録者）  
電子カルテデータと簡単に連携できる

MCDRS を用いることにより、SS-MIX2 の標準ストレージより患者基本情報、臨床血液検査結果、また、SS-MIX2 拡張ストレージより心電図、心エコー、心臓カテーテル検査結果を抽出して MCDRS にデータを格納することが可能である。

さらに、自然言語処理のためのフレームワークの検討として、形態素解析器の検討および医学用語辞書の精度向上のための検討を行った。申請時には MeCab の利用を想定していたが、京都大学で開発された Juman++ の性能が高いことがわかり、その利用を検討した。実際に用いるためには、形態素解析の精度ばかりでなく、辞書データのメンテナンス性や、解析結果の取得のしやすさ、実装のしやすさ、速度、安定性などが要求される。検討した結果としては今のところ十分に利用可能と考えており、今後、各施設より取得した実カルテ記載データ、特に数値情報等が混在するテキストデータでの検証を行う。また、辞書については急性心筋梗塞、心不全に関する用語のチェックを継続して行っている。病院情報システムからのデータ抽出としては、国立循環器病研究センター病院では N 社の病院情報システムが導入されており、データウェアハウス (DWH) が導入されている。DWH にはいわゆる電子カルテ本体より SOAP 記載が転送されており、これらのデータは抽出プログラムを構築することにより、DWH のデータベースからサンプルデータを抽出して、形態素解析の検証を実施した。

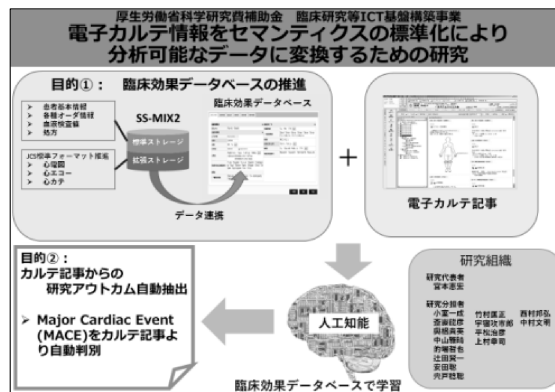
#### D. 考察

研究により、レセプト/DPC などの診療報酬請求情報を使用した分析、または厚生労働省電子的診療情報交換推進事業による標準化ストレージである SS-MIX2 データなどの電子カルテ情報を用いてビッグデータの分析において、MACE などのイベントをアウトカムにした研究をすることができるようになる。それにより高齢化社会の中で急増する循環器疾患の予後を改善させ医療費を適正化するための医療分析が可能となる。

#### E. 結論

この研究成果が電子カルテに実装されるようになれば、診療録の記載や各種検査・投薬オーダーを行う際の警告(アラート)の強化や、記載漏れや監査に対応した病名推薦システムを構築することができる。本研究によりレセプト/DPC などの診療報酬請求情報を使用した分析、または電子カルテ情報を用いてビッグデータの分析において、MACE などのイベントをアウトカムにした研究を実施することができるようになる。それにより高齢化社会の中で急増する循環器疾患の予後を改善させ医療費を適正化するための医療分析が可能となる。また、この研究成果が電子カルテに実装されるようになれば、診療録の記載や各種検査・投薬オーダーを行う際の警告(アラート)の強化や、記載漏れや監査に対応した病名推薦システムを構築することができる(図 3)。

図 3



循環器疾患の辞書作りは、電子カルテ記事の自然言語処理を行うのに必要な作業であり、他の研究にも応用活用であり、電子カルテに実装することが可能となる成果物である。心エコー、心臓カテーテル検査レポートを SS-MIX2 から取得できるようになれば、ベンダーの違う電子カルテ同士でもデータの連携が可能となる。

F. 健康危険情報  
なし

G. 研究発表

1. 論文発表   なし
2. 学会発表   なし

H. 知的財産権の出願登録状況

1. 特許取得       なし
2. 実用新案登録   なし
3. その他         なし