

政策科学総合研究事業（臨床研究等 ICT 基盤構築・人工知能実装研究事業）

分担研究報告書

研究課題：集中治療領域における生体情報や診療情報等を活用した人工知能（AI）の実装を推進するための基盤整備に係る社会的・技術的課題等についての実証的研究（19AC0201）

分担研究 2．集中治療室の時系列データの構造に係る調査

研究分担者

大嶽 浩司，橋本 悟，宮田 裕章，澤 智博，濱上 知樹

研究要旨

集中治療室で収集している診療データ（経過記録，観察指示，採血項目，指示等）は，各大学，企業毎に仕様・構造が異なっている．本研究ではその一部のサンプルデータを各大学，各企業から収集し，構造の課題の整理を行って，これらを元に，データクリーニング，個人情報秘匿化などを施されてデータ利活用が可能となる共通のデータベースの設計図を構築する．

研究目的

実際に臨床運用されている遠隔 ICU システムに蓄積された診療データを用いて，医療ユースケースに特化した機械学習技術の開発を行う．医療ユースケースを研究ビークルとして，説明性の高い機械学習技術の実現させるため，急性期医療に直接的に役立つドメイン固有機能の作り込みと，集中治療専門医に自動判断結果を理解し納得して受け入れてもらえるための説明機能を開発する．

遠隔 ICU システムには複数患者を効果的に管理するため，モニタリングと監視体制を構築した上で，重症度により患者をトリアージする機能が必要である．（次ページ図 1）遠隔 ICU を構築している病院群においては，これらのデータ収集環境および医師への情報提示インターフェースの構築を導入する環境が整っている．そのため，重症化予兆検知モデルに基づいたアラート発報およびその理由提示を，医療現場で実際に試す事が可能で，集中治療専門医のフィードバックもすぐに得られる環境を我々は有しており，研究遂行上の優位化が図れると考える．

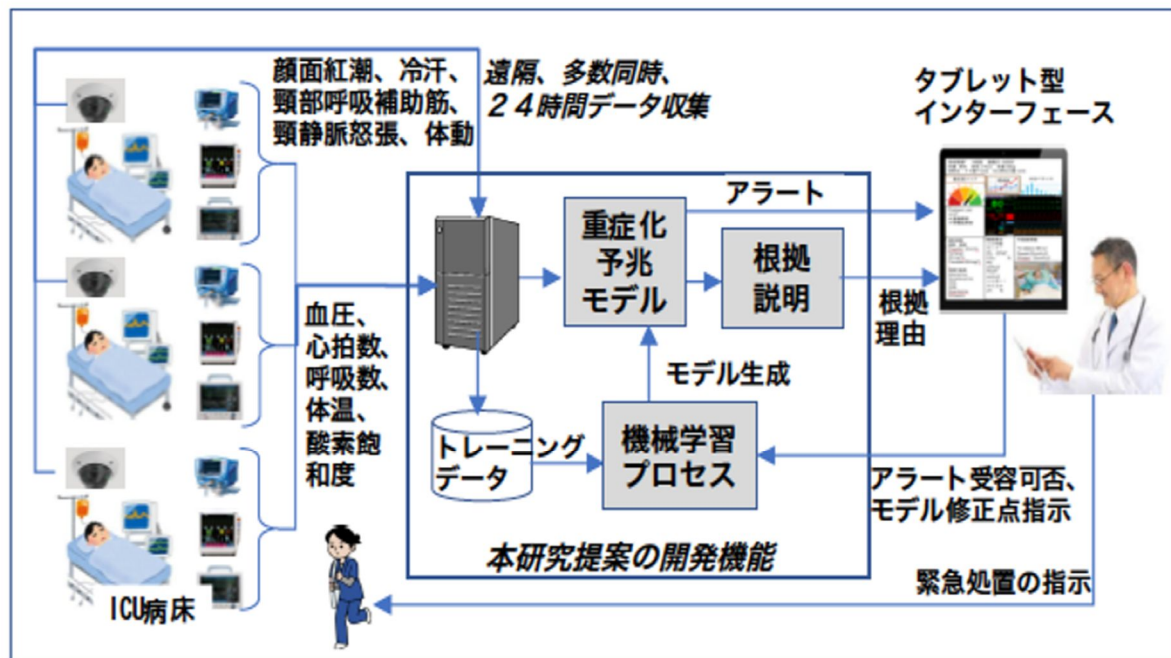


図 1 遠隔 ICU システムによる診療支援システムのイメージ図

ICU における重症度アルゴリズムを構築する上でのユースケースを設定する必要がある。疾患毎にアウトカム設定をおき，時系列パネルデータを用いた重症度アルゴリズムのパイロットモデルの構築を行う。

パネルデータとして ICU 病床で計測する患者のバイタルサインを用いると同時に，オープンデータ（MIMIC III）の活用も図る。

さらに，現状の患者背景を主とした日本 ICU 患者データベース（JIPAD (Japanese Intensive care Patient Database)）と本研究の時系列データベースを融合していく事で，データ利活用の幅が広がると思われる。通常の AI 解析に用いるデータは大規模な NCD のようなデータベースであるが，実臨床の場でリアルタイムに診療支援に用いるためにはパネルデータの利活用が欠かせない。重症度予測モデルを用いて，急性期現場の医療支援を行うことを本研究の目的とする。

・ All Japan ICU データベースとデータ利活用のシナリオ

各施設から収集したデータをクリーニングや個人情報の秘匿化などを施したのち，構造化されたデータベースを構築する。データ利活用に繋げるためのオープンなデータベース構築に向けた設計を行う。またこれらの構造化されたデータベースは将来的には，世界の ICU レジストリーデータと比較を行ったり，データ利活用の世界大会ともいえる Datathon などと連携をしたりすることを目指す。

個々の医療機関における個別患者の診療が，世界のビッグデータと繋がり，フィードバックをうけるようなダイナミックな医療の光景を視野に入れ，本研究が実施されている．

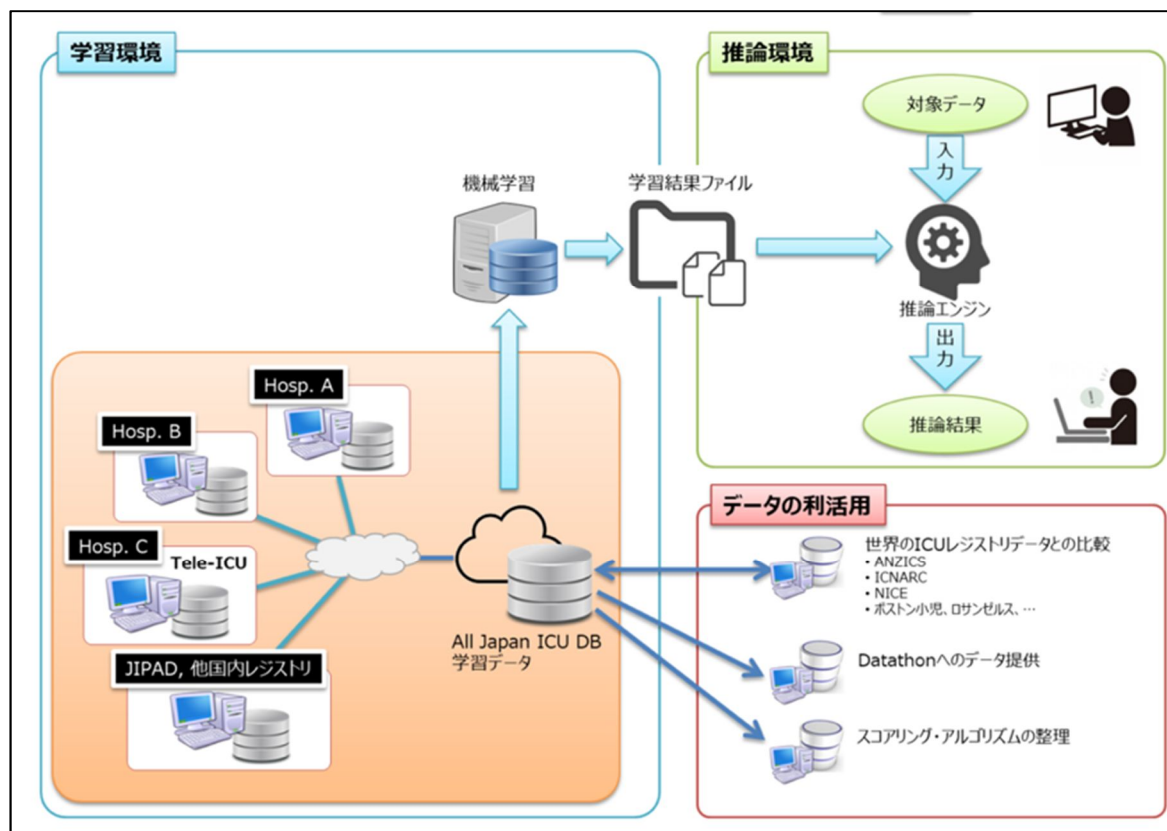


図2 遠隔 ICU システム，All Japan ICU データベースによるデータ利活用のイメージ図

データ構造，抽出項目等々の標準化がされた後には，各 ICU システムが導入されている施設からデータを抽出し，データベース構築を行うための仕組みについての調査を実施した．

データ収集においては，電子カルテは外部ネットワークと連携していないため，院内において ETL (Extract, Transform, Load) のシステムを開発する必要があり，外部のアップロードサーバの手前の中間サーバにおいて匿名化処理を行わねばならない．これらのデータフローに関して，効果的にデータを収集する運用フローをどの様に決めていくかについても調査を実施した．

ICU のデータベースから院内ネットワークを介して，ICU 内の中間サーバへデータを移行し，中間サーバにおいて匿名化処理を行ったのち，インターネットを介して ICU 標準化 DB へとデータをアップロードする方式について解析をした．これらを用いて，実際に各施設のデータを収集したデータベースにおいて，データの利活用を行える環境の醸成をめざす．

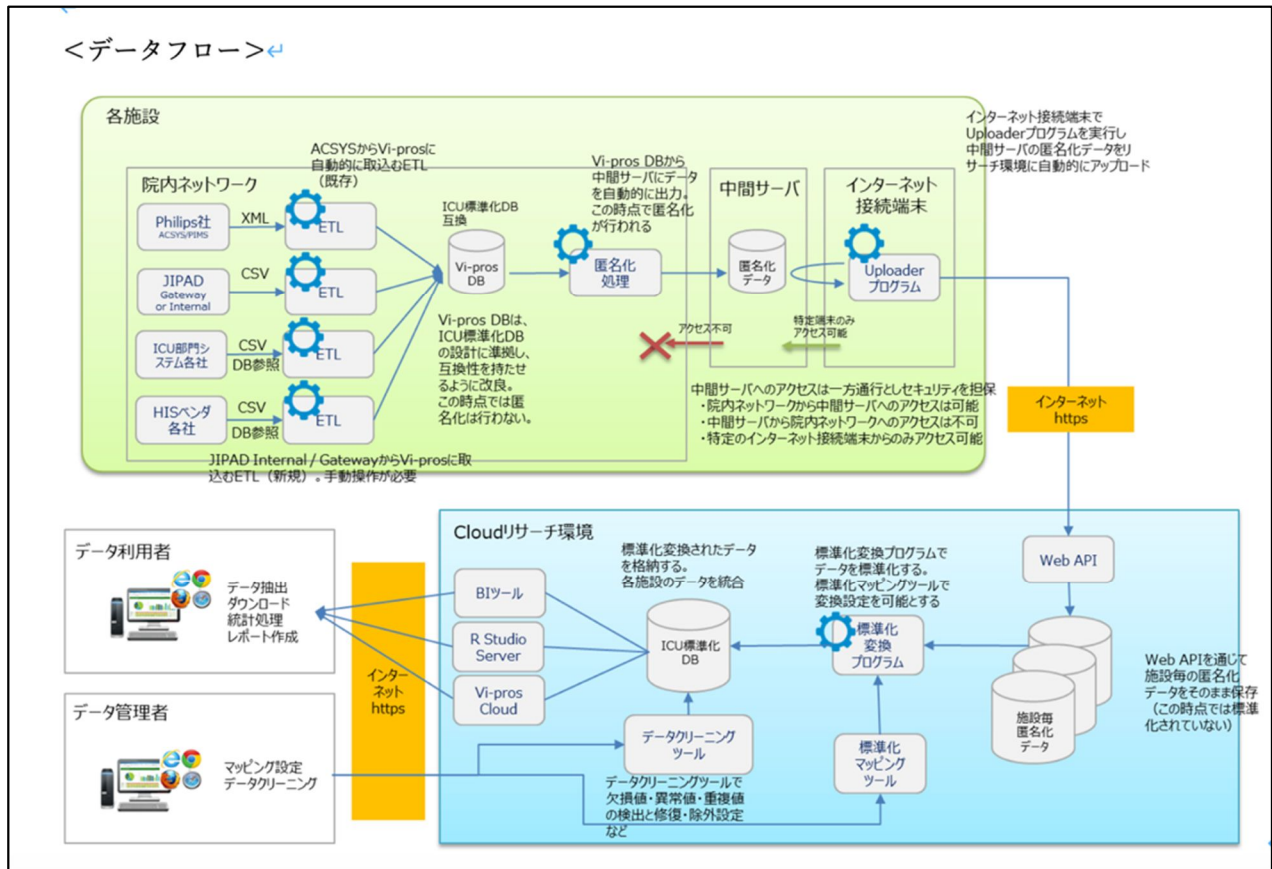


図3 ICUデータの抽出フロー

研究結果

集中治療室で収集している診療データ（経過記録，観察指示，採血項目，薬剤指示等々）を解析したところ，各大学，企業毎にその業務体系の違いなどのため抽出されたテキストやデータは仕様・構造が異なっていた．本研究ではその一部のサンプルデータを各大学，各企業から収集し，いかにマッピング手法やテキスト構造化手法を駆使して単一のプラットフォーム上で共通データの蓄積が行えるかを検討していくための整理を行った．

・多パラメータ診療データに対する人工知能の有効性の検討

集中治療室における診療データは非常に複雑かつ，多岐にわたるため，先行研究として患者ごとのばらつきが一定の範囲内でおさまり，パラメータの数も少ない昭和大学の病院で行われた手術中患者のバイタルサインを用いて，横浜国立大学にて人工知能を用いたイベント予測を行った．

術中バイタルデータは，複雑な非線形を有する多次元時系列データである．本研究担当では，多次元時系列データからの特徴抽出，現象予測，イベント予測に特化した人工知能アルゴリズムの開発に着手した．特に今期は，強化学習に基づく2つの機械学習アプローチを試行し，それぞれの特徴と性能を評価した．1つめは異常とログデータの因果関係を推定するためにTD(Temporal Difference)学習でネットワークの最適化を行うことで，異

常の期待値を価値関数として近似する手法を新たに開発した。価値関数の遷移に注目することで、どの状態が異常やイベントに影響しているかを可視化することが可能になった。また、価値関数のモデルに Attention 機構を導入することで、状態中のどの部分に注目しているかが可能になった。2つめは、事前知識が不要な深層生成モデルを用いたデータ拡張による汎化性能向上を目的とした。多次元時系列データを2次元画像情報に埋め込み、この画像を PGGAN (Progressive Growing of GANs) で生成する手法を開発した。医師がモニターの画面情報から直観的に状態を読み取り短期での予測をするプロセスを模擬していることに相当する。疑似データおよび実データによる検証の結果、基本的な性能が確認された。さらに、追実験により特徴マップ数のチューニングによる汎化性能の大幅な向上が確認された。最終的に、12分以内のイベント予測について実用性が示された。

・ICUにおけるデータ抽出方式の選択

データ構造、抽出項目等々の標準化を視野に入れ、どのような方式で ICU データを抽出すべきか実証するため、横浜市大および、京都府立医大の ICU システムからデータのパラメータ抽出を行い、構造化を行うための出力されたデータ方式の解析および効果的なデータ収集に必要な運用フローを検討した。

ICU のデータベースから院内ネットワークを介して、ICU 内の中間サーバへデータを移行し、中間サーバにおいて匿名化処理を行ったのち、インターネットを介して ICU 標準化 DB へとデータをアップロードする方式は、施設及び電子カルテが違っていても、中間サーバでデータ処理を行えば、一定程度可能であることが検証された。ただ、さらに多くの施設のデータを集積するあるいは刻々とアップデートされる現場のモダリティに対応していくには、データ形式の規格だけでなく、接続・運用・データプロセスまでを含めた標準化を行う必要があることが実感できた。今後は、各施設のデータを収集したデータベースにおいて、データの利活用を行える環境の醸成の方法の検討を重ねる。

・医療情報標準規格の視点での検討

集中治療医学分野での活用を想定される医療情報標準規格について、用語・コード、電文・通信、文書・画像、モデル・構造の別に各規格を整理した。また、厚生労働省標準規格において集中治療医学に於ける適用可能性について検討した。保健医療福祉情報システム工業会 (JAHIS) にて提示されている標準規格について集中治療医学分野での適用候補となる可能性について調査を実施した。

・システム基盤についての検討

主にハードウェアの側面から集中治療医学分野における医療情報システムのアーキテクチャについて検討した。サーバー系のインフラについては、従来のクライアント・サーバー型での運用形態に加えて、クラウドシステムやハ

イブリッドクラウドシステムでの運用を視野に入れ，仮想化サーバーインフラおよびハイパーコンバージドインフラの現状について調査を実施した．

・アプリケーションアーキテクチャについての検討

現在市販されている集中治療医学分野関連のアプリケーションの殆どはクライアント・サーバー型アーキテクチャを採用しており，旧来の Windows exe 型アプリケーションとソケット通信を組み合わせる形態となっている．一方で，一般社会に普及しているアプリケーションおよび米国等の先進事例が普及を目指している医療アプリケーションにおいては，ウェブ型で REST API (REpresentational State Transfer Application Programming Interface) が採用されており，日本の医療情報システムにおいても採用可能性が検討されている．HL7 FHIR(Fast Healthcare Interoperability Resources)の調査を含め，アプリケーションアーキテクチャについて検討を行った．

考察

日本 ICU 患者データベース (JIPAD (Japanese Intensive care Patient Database))では現在全国の 60 施設以上の特定集中治療室から集まった 13 万症例を超える症例が集積されている．これらのデータを使って，諸外国の ICU とのパフォーマンスの比較や各施設の実死亡率・予測死亡率比を算出してベンチマークとして使用しているが，これらのデータに加えて本研究で得られるデータを組み合わせることでより精度の高い，また AI 予測などが可能なデータベースが構築できると予想される．そのためには，ICU データの標準化が必要である．現状，ICU においては JIPAD で取り扱っていない下記のデータの標準化を定義する患者情報・入退院情報・ICU 入退室情報・主病名，副病名・基礎疾患・慢性疾患・人工呼吸器指示，実測・特殊治療指示，実測・バイタル情報実測値，確定値・検査結果・注射指示，実施・IN/OUT 情報・スコア (看護必要度，APACHE-II，APACHE-III，PIM2，PIM3，SOFA，DIC，SAPS2...)・経過記録，サマリ等がある．

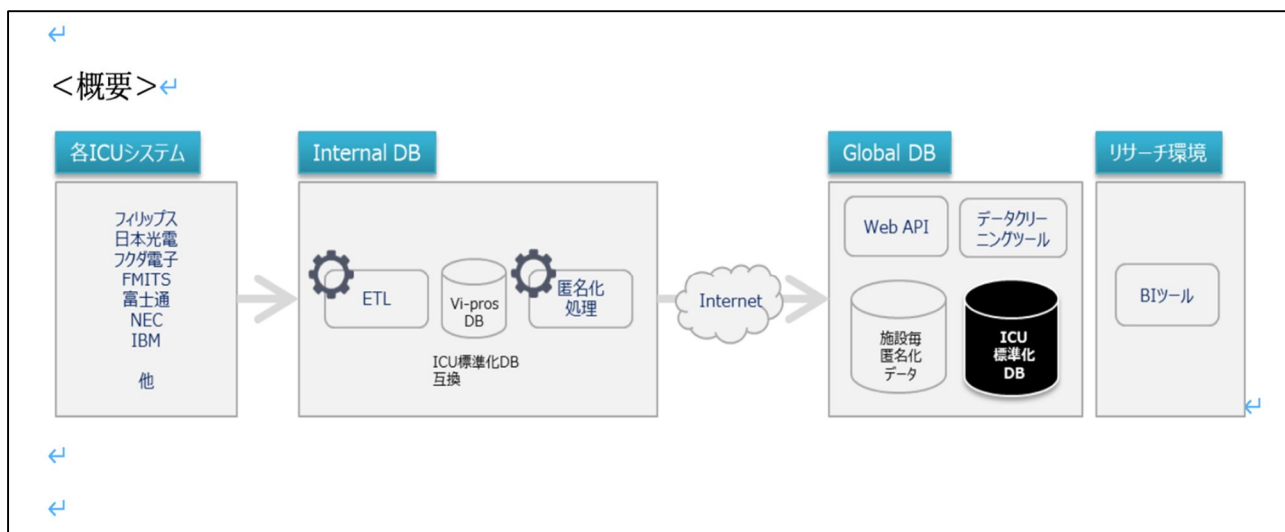


図4 データベース構築および利活用環境の醸成の概要図

これらのデータを多くの施設から抽出する事で、データ構造化の差異に関する調査を引き続き行う。標準化の調査とともに、実際のデータ収集のシステムも並行して開発していく。

また、データ利活用の現状の課題について整理することができたため、これらの課題を今後の研究課題として解決策の提案を進めていく必要があると考えている。こうしたデータ利活用をする際の重要な視点として、アウトカム設定がある。何のためにデータ利活用をするのかという点を突き詰めていく必要がある。本研究の目的は集中治療室における時系列データを用いたAI解析のアルゴリズムを構築する事である。このアルゴリズムを作る事で、患者の予後の質の改善へ結びつけ方や働き方改革へ影響などと言った視点で今後の研究を進めていく必要があり、本年度の成果は今後の研究の方針に活用できると考えている。