



私のブックマーク

医療情報処理—病院のカルテデータを中心に—^{†1}

香川 璃奈 (筑波大学医学医療系)

1. はじめに：医療情報学の全体像と医療情報処理

医療現場では患者と医療従事者のそれぞれが情報処理を行い、実務の背景では自治体や国などの行政関係者が適切な制度設計のための情報処理を日々繰り返しています。例えば医者は、患者の診察や検査によって得られた情報をリアルタイムに統合して、次にどのような診察や検査を優先すべきか（ときに患者をいかに説得するか）を決定して次の診察・検査に進みます。この情報収集と意思決定の繰り返しにより得た情報を統合して診断を行い治療が進み、患者が元の健康な状態に戻るまで情報収集と意思決定が繰り返されます。

医療情報学 (medical informatics) では、このような医療に関するあらゆる情報処理過程のメカニズムを明らかにし、それを効率的に実応用する手法を開発し、加えて、研究開発の基盤となる国際標準化事業 (ISO [1], 欧州標準化委員会 [2] などが先導する)、地域や時代に応じたシステム開発や制度設計を行います。例えば途上国向けのシステム開発 [3] や人工知能・機械学習に基づく医療機器規制についての提言 [4] も行います。

医療情報学の対象は多岐にわたりますが、研究テーマを絞る観点からは、データの発生元を医療施設 (病院または診療所)、行政、患者、そして基礎医学研究に整理でき、また研究の応用先も同様に、医療施設および医療従事者、行政、患者、基礎医学研究に整理できます。本稿では特に、医療施設において日々の臨床業務で発生してカルテに蓄積されたデータを、医療従事者への応用を念頭に解析する研究を医療情報処理と呼ぶことにします。電子化されていないカルテの議論は省略します。

2. 医療情報処理の研究リソース

2.1 主要なデータセット

カルテデータは、形式も臨床現場での用途も異なる多様なデータから構成されます。データ形式に基づき整理すると、検査結果や処方日数などの数値、心電図や脳波などの波形、CTや内視鏡などの画像、医療従事者の情報収集と意思決定の繰り返しの詳細が記録される自然文テキストまで含まれます。研究利用するデータセットを検討する際には、臨床現場での用途の観点から、臨床情報と、会計データである請求情報の二つに大別できます。本邦からの申請手続きが比較的簡便かつ利用費用が不要または少額なデータセットに絞ってご紹介します。というのも、カルテのデータセットは、プライバシーの観点から利用者や利用場所が制限される場合や、仮名化 (匿名化) 処理にコストがかかるために利用料が必要な場合があるのです。この先、昨今の画像処理研究の発展のおかげで多くの資料が存在する医療画像、およびバイオインフォマティクス分野に詳しい遺伝子やタンパク質などに限定された内容は割愛します。

§1 臨床情報

診断名や検査値などの数値データが格納されている米国の ICU (集中治療室) データの e-ICU Collaborative Research Data [5] および、それらに加えて少数のテキストを含む MIMIC-III [6] の利用には Web 上での倫理講習受講と第三者への照会が必要です。英国の診療所から検査値や処方内容などを収集した CPRD [7] は、研究計画が承認されれば購入できます。ほかにも米国の心電図などの波形データが PhysioNet [8] に公開されています。日本語のカルテテキストでは擬似診療録データ [9] が提供されています。また、本邦では法律に基づき、がんと診断されたすべてののがんに関するデータが全国がん登録 [10] として収集されており、申請が承認されれば利用できます。その他、政府統計 [11] には本邦の病院数や電子カルテの普及率などが公開されています。通所・訪問リハビリテーションに関

^{†1} http://www.ai-gakkai.or.jp/my-bookmark_vol34-no6

する VISIT [12] など、研究利用を志向して構築中のデータセットも存在します。

§2 請求情報

医療費は、患者が一部を医療機関に支払い、場合によっては公費からの支出があり、残りは保険者（例：全国健康保険協会など）が医療機関に支払います。実際に行われた医療行為に基づいて医療費が決定されるため、医療機関から保険者への請求情報には、処方内容や実施された処置（例：透析、人工呼吸など）などが含まれます。このデータには検査結果や治療のアウトカム（症状がどの程度改善したか、疾患は治癒したか、など）が含まれない欠点はあるものの、規模が大きいことが利点で、研究利用の体制は整っています。本邦では保険診療のほぼ全症例が含まれるレセプト情報・特定健診など情報データベース（NDB）[13]を申請に基づき利用でき、40～74歳に行われるメタボリックシンドロームの予防と改善のための特定健診・特定保健指導情報も併せて利用可能です。これらの一部の集計結果はNDBオープンデータ[14]として公開されています。解析を行ううえで、請求情報に含まれる診断名が患者の病態を必ずしも反映していない問題[15]は世界的に知られています。一部の病院の入院患者に限定されるものの患者の真の病態を反映した病名の割合が高いDPC[16]や、要介護認定情報や介護報酬がまとまった介護DB[17]も申請手続を経て利用できます。全米の病院からのランダムサンプリングに基づいた入院患者（NIS）[18]、救急（NEDS）[19]、小児（KID）[20]の診療報酬データ、およびこれらの州別データもWeb上の講習を受講したうえで契約を交わすことで購入可能です。

2.2 用語集・コード集など

それぞれが異なる目的に沿って作成されているため、研究ごとに適切な選択が必要です。

・病名、症状など

ツリー構造を有する国際疾病分類（ICD [21]）が世界保健機構（WHO）により定められており、本邦ではICDに準拠した標準病名マスタ[22]が整備されています。しかしICDはあくまで死因分類を基盤とするため、ツリー構造と医療従事者がつく知識構造が必ずしも一致しません。例えば、同じ臓器に生じる疾患のうち腫瘍だけ別の大分類下であることや、同じ症状を呈する疾患どうしが異なる分類下に存在することが研究目的によっては問題になり得ます。その他、医薬品の副作用が充実したMedDRA[23]、処置や社会的内容など病名以外の計24カテゴリーが網羅されたSNOMED-CT[24]（年間数億円を超えるライセンス料のため本邦では閲覧しかできない）、これらをひも付けるUMLS[25]などが利用されており、日本語で疾患の発症から治癒までの機序の記述を試みる臨床医学オントロジー[26]も開発中です。Human Phenotype Ontology[27]は研究利用を志向するため、一般的な臨床現場では考慮しない用語も多く含みます。

・薬剤

みなさんがご自身やご家族の内服薬の名前を覚えているとしたら、それは商品名でしょう。薬剤には、後発（ジェネリック）医薬品など商品名が異なっても成分は同じ場合があります。さらに成分が異なっても薬効（薬の効きめのこと。血圧を下げる、血糖値を下げる、など）は同じものもあります。研究目的に応じて対象薬剤を分類する際に有用なのが、薬剤コードです。薬剤コードは本邦[28]のみならず世界中に多くの種類が存在[29]し、コード間のマッピングが課題になっています。医療従事者が臨床現場で考える薬効分類と同等のツリー構造がわかりやすいと筆者が感じるのがWHOが定めるATCコード[30]、諸外国のデータセットで利用されることが多い印象をもつのがRxNorm[31]です。

・検査

一言で尿検査といっても検査される項目は多数ありますし、また「赤血球数の検査」や「白血球数の検査」は採血（静脈血）検査でも尿検査でも行われます。赤血球数や白血球数といった検査項目と、静脈血や尿などの検査材料の組合せを一意に定めるために、本邦ではJLAC 10[32]という現行コードに代わるJLAC 11を開発中です。米国ではSNOMED-CTとも対応がとれるLOINC[33]が主に利用されています。

・その他

- ・ FMA [34]：解剖学オントロジー
- ・ Comejisyo [35]：医療領域全般の日本語用語の辞書。看護領域の形態素解析目的で開発されています。
- ・ 国際生活機能分類（ICF）[36]：衣服を着ること、ジェスチャの理解、といった健康（生活機能）に関する用語がWHOにより整理されています。
- ・ OMOP-CDM [37]：病態から薬剤、検査結果まで幅広く表現する共通用語。利用が世界的に広がり、日本語対応に向けた動きも始まっています。
- ・ 手術基幹コード（STEM7）[38]：乳腺の悪性腫瘍の切除といった医療行為の日本語コードです。WHOは

医療行為の国際分類 (ICHI) [39] を開発中です。

2.3 教科書など

§1 医療情報学について

研究テーマを俯瞰する目的では、Shortliffe 監修教科書 [40]、国際医療情報学会が毎年公開する *Yearbook of Medical Informatics* [41]、米国医療情報学会による *Online On-Demand Learning* [42] (有料)、オンラインの *journal club* [43] (有料) を利用できます。米国内では複数大学での講義シリーズ *AMIA 10x10* [44] を受講できます。実際の医療データの例やリソースの詳細を網羅的に知る目的では荒牧の著書 [45] が有用です。

§2 医療・医学について

疾患や治療の概要を初めて学ぶためにはみえるシリーズ [46] やイヤートート [47] が有用です。より詳細な各疾患の診断基準や最新の治療方針は自国の診療ガイドライン [48] の参照が効率的です。ただし、ガイドラインは医療現場に強制力をもたず、一部の分野や疾患では作成が追いついていません。臨床現場における医療従事者の情報処理過程が垣間見える一般書として *How Doctors Think* [49] は誠実な内容だと感じています。

3. 主な研究課題

3.1 応用志向の研究課題

§1 診療支援・自動診断

診療支援研究は、知る限りで 1950 年代に提案された医師に診断候補を示す *logoscope* [50] (写真 [51]) に始まり、1970 年代にはエキスパートシステムに基づき抗菌薬の推薦システム *MYCIN* [52] が開発されました。現在でも最も重要な課題であり、医療者を補助する目的での検査や治療や診断候補の推薦 [53]、特に人手不足が著名な専門領域や地域において医療者を代替する目的での自動診断 [54] などの研究が行われています。医療従事者を目指す学生の教育 [55] も広い意味での診療支援といえるでしょう。

§2 患者情報の識別

カルテにおいて、患者の病名、喫煙や飲酒の情報、主観的な訴え、日々の症状の変化などは、構造化されて入力されることは少なく、医療従事者がテキストに自由に記載します。しかし多忙な臨床現場においてテキストは丁寧には記載されません。そこで、テキストに基づく上記の情報の抽出 [56] や関係抽出 [57] の研究が行われています。また多様なカルテデータの組合せに基づいてある特定の疾患や症状を有する患者集団をいかに識別 [58] するかという課題も解決されていません。ある程度定型化された記載も存在しますが、患者の叙述的な訴えや、現場の医師でも判断に困っている状況での記録など、そもそも (半) 構造化入力を促すことが難しい情報も多いため、筆者は、臨床現場に (半) 構造化入力を強制すべきとも限らないという立場をとっています。

3.2 基礎的な研究課題

§1 データの特性に基づく課題

入院したときは毎日採血を行った患者でも、元気になって退院したあとは月に 1 回の外来受診に合わせて採血を行うでしょう。毎月真面目に外来受診する患者でも大雪のせいで止むを得ず採血のタイミングがずれることもあります。特に時系列データを解析する際には、不規則に生じるデータの欠損を補完するだけでなく、その特徴を踏まえたモデル構築 [59] が行われています。なお、ICU では患者の血圧や体温などは機械で定期的に測定し続けることが多く、欠損が比較的少ない時系列データを入手できます。また、教師データ作成時に複数作業員間で結果が異なる報告 [60] は医療情報処理でも認められており、教師データにノイズがある場合の研究 [61] も行われています。

§2 社会的特性に基づく課題

医療情報処理で扱う病気のデータは患者の *sensitive data* (機微情報) です。プライバシー保護を目的とした、*de-identification* [62] や擬似データ生成 [63] の手法開発の必要性は高いままでしょう。また、医療従事者から患者への説明不足や誤った判断に基づく医療行為が訴訟に直結するという、臨床現場の社会的な特性を反映して、*interpretability* [64]、*robustness* [65]、*fairness* [66] を満たす機械学習手法の応用は今後ますます盛んになると予測されます。

3.3 医療従事者の情報処理に関する研究課題

医療情報処理を深めるための真の基盤となる、医療従事者の情報処理とその特性を理解することを目的とする研究は、1965年に心理学者が医師の診断過程をツリー構造で表せると明らかにした研究が始まりといわれています [67]。しかし、この課題は研究が盛んとはいえません。研究に必要な量と質のデータや実験被験者となる医療従事者を集めることの難しさが一因と考えられます。近年行われた研究では、カルテテキスト内の病名などの半構造化記載の医師ごとの相違 [68]、診療支援システム開発が成功する人的要因 [69]、カルテの電子化により失われたとされる情報 [70] などが明らかにされています。

3.4 研究開発の基盤となるトピック

§1 電子カルテの普及

国民皆保険制度を採用する本邦では、1960年代から請求情報を電子化するシステムの導入が進みました。その後、e-文書法 [71] に先駆けて1999年に電子カルテが正式なカルテとして国から認められ [72] ました。現在は大規模な病院の85%以上で電子カルテが導入されていますが、地域や病院規模による差 [73] が認められます。なお、国や地域によっては、患者の診察データの保管責任が患者本人にありますが、本邦の現状では医療法 [74] の規定により医療機関に一定期間の保管義務があります。

米国では2009年のHITECH法に基づく政府の奨励金により、診療所で40%程度であった電子カルテの普及率が2017年までで約85%まで急速に進みました [75]。2015年時点の各国の状況はWHOが公開 [76] しています。

§2 情報交換標準規格

医療情報処理の発展のためには、電子カルテの情報を多施設間で交換することが必要です。最も関連が深い情報交換標準規格の一つがHealth Level Seven (HL7) [77] です。HL7協会により開発された標準規格で、昨今急速に導入が進んでいるのがFHIR [78] という短期間での開発・導入を目的として策定されたものです。KDD 2019のオープニングキーノート [79] でも言及され、SMART on FHIR [80] というアプリ開発用プラットフォームも米国を中心に普及しています。

本邦ではHL7標準の中でも、より詳細な標準に基づくSS-MIX2 [81] が厚生労働省標準規格 [82] として定められ全国1000施設以上の電子カルテで導入され、全国規模の医療データ収集事業 [83] などで利用されています。しかし、すべての電子カルテ製品がこれに準拠しているわけではありません [84]。

§3 プライバシー

医療情報処理においては、患者、場合によってはデータ作成に関与した医療従事者をも保護する必要があります。私の治療データがこんな研究に使われるとは思ってもいなかった！ という苦情を後から受けないように、研究者自身ひいては研究コミュニティを守るためにも必要なだご理解ください。本邦では人を対象とする医学系研究に関する倫理指針 [85] に従いますが、研究によって別指針に基づく場合がありますので、厚労省の指針 [86] をご確認ください。他国で研究を行う場合にはその国の法律に則らなければなりません。

プライバシーの観点から、データセットやコーパスの公開が困難であることは、人工知能分野の急速な発展の中で医療情報処理が出遅れる一因となっています。本邦では通称次世代医療基盤法 [87] の施行によりデータ利用の促進が期待されていますが、運用はこれからでありその実力は未知数です。米国はHIPAA [88] により医療情報における個人情報に明確に定義されながら公開されているデータはごく一部であることが、プライバシーの難しさを露呈しているともいえます。その一方で、過度な自己防衛で研究者の利便性を損なう必要もありません。例えば本邦では、通称3省3ガイドライン（厚労省 [89]、経産省 [90]、総務省 [91]）に準拠しているクラウドサーバには医療データを置くことができます。

4. まとめに代えて：COIおよび倫理的配慮

医療情報処理を含む医療系研究では、些細な発表にも必ずCOI（利益相反） [92] や倫理的配慮 [93] を明記することが当然の責務として根付いてきました。人工知能研究の研究者の皆様には驚かれることもある文化の違いですが、けっして研究の自由を制限するものではありません。本稿に関連するCOIはありません。

謝 辞

執筆に際して有意義なコメントを下された本武陽一さん(統計数理研究所), 三谷知広さん(東京大学大学院)にお礼申し上げます。本稿は多くの先生方との議論の蓄積なしには書けませんでした。最後に, 貴重な機会を与えてくださった人工知能学会誌編集委員会に心より感謝申し上げます。

- [1] <https://www.iso.org/committee/54960.html>
- [2] <https://www.cen.eu>
- [3] <https://imia-medinfo.org/wp/health-informatics-for-development/>
- [4] https://www.amia.org/sites/default/files/AMIA-Response-to-FDA-AIML-SaMD-Modifications-Draft-Framework_0.pdf
- [5] <https://eicu-crd.mit.edu/>
- [6] <https://mimic.physionet.org/>
- [7] <https://cprd.com/home>
- [8] <https://physionet.org/physiobank/>
- [9] <https://www.gsk.or.jp/catalog/gsk2012-d>
- [10] https://ganjoho.jp/reg_stat/can_reg/national/datause/general.html
- [11] https://www.e-stat.go.jp/stat-search?page=1&bunya_l=15
- [12] https://www.mhlw.go.jp/file/05-Shingikai-12301000-Roukenkyoku-Soumuka/0000180922_5.pdf
- [13] https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryuu/iryuuhoken/reseputo/index.html
- [14] <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000177182.html>
- [15] <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0104519>
- [16] https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryuu/iryuuhoken/dpc/index.html
- [17] https://www.mhlw.go.jp/stf/shingi2/0000198094_00002.html
- [18] <https://www.hcup-us.ahrq.gov/nisoverview.jsp>
- [19] <https://www.hcup-us.ahrq.gov/nedsoverview.jsp>
- [20] <https://www.hcup-us.ahrq.gov/kidoverview.jsp>
- [21] <http://www.byomei.org>
- [22] <http://www2.medis.or.jp/stdcd/byomei/index.html>
- [23] <https://www.meddra.org/how-to-use/support-documentation/japanese>
- [24] <https://www.nlm.nih.gov/healthit/snomedct/index.html>
- [25] <https://www.nlm.nih.gov/research/umls/index.html>
- [26] <http://www.m.u-tokyo.ac.jp/medinfo/medont2009proj/>
- [27] <https://hpo.jax.org/>
- [28] <https://www.data-index.co.jp/knowledge/detail1-1.html>
- [29] <https://forums.ohdsi.org/t/drug-concept-navigation-and-source-codes/2035>
- [30] https://www.kegg.jp/kegg-bin/get_htext?jp08303.keg
- [31] <https://www.nlm.nih.gov/research/umls/rxnorm/docs/rxnormfiles.html>
- [32] <https://www.jslm.org/committees/code/index.html>
- [33] <https://loinc.org/>
- [34] <https://bioportal.bioontology.org/ontologies/FMA>
- [35] <https://ja.osdn.net/projects/comedic/>
- [36] <https://apps.who.int/iris/bitstream/handle/10665/42407/9241545429-jpn.pdf?sequence=313&isAllowed=y>
- [37] <https://www.ohdsi.org/data-standardization/the-common-data-model/>
- [38] <http://www.gaihoren.jp/gaihoren/public/info/info-180420.html>

- [39] <https://mitel.dimi.uniud.it/ichi/docs/#introduction>
- [40] https://www.researchgate.net/publication/282506401_Biomedical_Informatics_Computer_Applications_in_Health_Care_and_Biomedicine
- [41] <https://www.thieme-connect.com/products/ejournals/journal/10.1055/s-00034612>
- [42] <https://www.amia.org/online-demand-learning>
- [43] <https://www.amia.org/category/webinars/journal-club>
- [44] <https://www.amia.org/amia10x10>
- [45] <http://www.coronasha.co.jp/np/isbn/9784339027624/>
- [46] <https://www.byomie.com/>
- [47] <https://www.yearnote.com/>
- [48] <https://minds.jcqhc.or.jp/>
- [49] https://en.wikipedia.org/wiki/How_Doctors_Think
- [50] [https://doi.org/10.1016/s0140-6736\(54\)91437-3](https://doi.org/10.1016/s0140-6736(54)91437-3)
- [51] <http://broughttolife.sciencemuseum.org.uk/broughttolife/objects/display?id=93319>
- [52] [https://doi.org/10.1016/0010-4809\(75\)90009-9](https://doi.org/10.1016/0010-4809(75)90009-9)
- [53] <https://arxiv.org/pdf/1906.00346.pdf>
- [54] <https://www.cnn.com/2019/09/12/ges-health-unit-wins-first-fda-clearance-for-ai-powered-x-ray-system.html>
- [55] <https://doi.org/10.2196/jmir.5461>
- [56] <https://doi.org/10.1093/jamia/ocy173>
- [57] <https://doi.org/10.1186/s12911-019-0769-0>
- [58] <https://www.phekb.org/>
- [59] <https://doi.org/10.1016/j.artmed.2013.01.003>
- [60] <https://doi.org/10.1016/j.ijmedinf.2016.09.014>
- [61] <https://arxiv.org/abs/1411.7718>
- [62] <https://doi.org/10.1093/jamia/ocw156>
- [63] <https://doi.org/10.3233/SHTI190249>
- [64] <https://arxiv.org/pdf/1907.07374.pdf>
- [65] <https://arxiv.org/pdf/1804.00504.pdf>
- [66] <https://doi.org/10.1371/journal.pmed.1002689>
- [67] https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_uri&item_id=6850&file_id=1&file_no=1
- [68] <https://doi.org/10.1093/jamia/ocv211>
- [69] <https://doi.org/10.1016/j.ijmedinf.2016.12.001>
- [70] <https://doi.org/10.1016/j.ijmedinf.2015.09.004>
- [71] https://www.mhlw.go.jp/file/05-Shingikai-12601000-Seisakutoukatsukan-Sanjikanshitsu_Shakaihoshoutantou/0000166260.pdf
- [72] https://www.mhlw.go.jp/www1/houdou/1104/h0423-1_10.html
- [73] https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450021&bunya_1=15&tstat=000001030908&cycle=7&tclass1=000001123595&tclass2=000001123598&stat_infid=000031780693
- [74] https://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=323AC0000000205
- [75] <https://dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php>
- [76] http://apps.who.int/iris/bitstream/10665/204523/1/9789241565219_eng.pdf
- [77] <http://www.hl7.jp/>
- [78] <https://www.hl7.org/fhir/>
- [79] <https://www.youtube.com/watch?v=skqyBA1MxXQ>
- [80] <https://apps.smarthealthit.org/>
- [81] <http://www.jami.jp/jamistd/ssmix2.php>

- [82] <http://helics.umin.ac.jp/helicsStdList.html>
- [83] <http://jdreams.jp/system/>
- [84] http://www.ss-mix.org/cons/product_car01.html
- [85] <https://www.mhlw.go.jp/file/06-Seisakujouhou-12600000-Seisakutoukatsukan/0000168764.pdf>
- [86] <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hokabunya/kenkyujigyou/i-kenkyu/index.html>
- [87] https://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=429AC0000000028
- [88] https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf?fbclid=IwAR3D9PHIU515NFoEzr90RJMJlaXBYaP0B_7FiBmSrZfg3c24JAJ1oCHpDW8
- [89] <https://www.mhlw.go.jp/stf/shingi2/0000166275.html>
- [90] https://www.meti.go.jp/policy/it_policy/privacy/iryoug1v2.pdf
- [91] http://www.soumu.go.jp/menu_news/s-news/01ryutsu02_02000209.html
- [92] http://jams.med.or.jp/guideline/coi-management_qa03.html
- [93] https://www.nurse.or.jp/nursing/education/gakkai/happyo/pdf/50_rinri_kijyutu.pdf