

「レセプト情報をAIで類型化することによる医療費の分析及び活用方策の検討のための研究」

分担研究報告書

「レセプト審査におけるAIの活用可能性に関する研究」

研究分担者 伊藤 善典 埼玉県立大学・保健医療福祉学部 教授

研究要旨

A. 研究目的・背景

後期高齢者医療等のレセプト件数が増加する中、国民負担の軽減等の観点から、都道府県国民健康保険団体連合会（国保連）のレセプト審査業務のあり方について関係者から様々な要請がなされている。本分担研究では、レセプト審査業務の効率化・高度化を推進する観点から、AIの活用の具体的なあり方、その有効性、活用を行うに当たって解決されるべき課題等について検討を行う。

B. 研究方法

まず、AIの具体的な活用方法として次の2つを採り上げ、その有効性等について検証を行った。使用するデータは、複数の国保連から匿名の審査済レセプトデータの提供を受けた。

① AIの技術を用いて生成された患者像の活用（分担研究1参照）

② 審査済レセプトを学習させたAIの活用

次に、検証結果を踏まえ、国保連におけるレセプト審査の業務フローを念頭に置きつつ、AIを実際にどのように活用できるのか、そのためにはどのような課題があるのか等について考察を行った。

C. 研究結果

まず、AIの技術を用いて生成された患者像に基づき、患者像から診療行為等の数量や回数がかい離しているレセプトを抽出した。その結果、かい離が大きければ査定につながるというわけではないが、最もかい離度が高いレセプトでは、査定率が突出して高いことがわかった。

次に、審査済レセプトを学習させたAIの活用について検証を行った。AIにレセプトの審査結果を多数学習させ、他のレセプトを用いて査定又は査定対象外の判定を行わせたところ、査定対象外となるべきレセプトや外来のレセプトについては、精度の高い判定が可能であることが確認された。また、AIが学習するデータ量が大きいほど、判定精度が向上することが確認された。

D. 考察

AIの技術を用いて生成された患者像については、直接的な審査への使用には未だ課題があるものの、重点的に審査を行う対象として、かい離度の高いレセプトやそれらの数が多い医療機関を抽出する際には有用であると考えられる。

また、レセプト審査業務にAIを活用することについては、次のように考えられる。

① AIの活用は、業務を効率化するうえで有効な手段となりうる。

② AIが学習するデータ量や項目を増加させる等により、判定精度を向上させることができる。

③ AIを活用する分野や業務については、全てのレセプトを対象とする方法だけでなく、判定精度の高い分野に絞って活用する方法もありうる。

④ AIに医学的判断を含む審査結果を学習させることにより、現在、事務職員が処理している業務を代替できる可能性がある。ただし、コンピュータチェックは、AIとは機能が異なるため、併用する必要がある。

⑤ AIの活用にあたっては、導入費用や軽減できる作業量等を使用した、費用対効果の検討が必要である。

⑥ AIを活用して行った審査事務共助の結果については、判定理由を審査委員及び事務職員に明示できるようにすることが求められる。

⑦ AIの技術を用いて生成された患者像を活用すれば、審査における地域差を明らかにすることができる可能性がある。

E. 結論

レセプト審査業務でAIを活用するにあたっては、判定精度の向上、費用対効果の検討などいくつかの課題があるものの、AIの活用範囲が広がれば、業務の効率化・高度化による国民負担の軽減や審査委員の負担軽減などメリットがあるため、関係者の理解を得つつ、活用に向けた検討を加速することが望まれる。

A. 研究目的、背景

2017年1月、厚生労働省の「データヘルス時代の質の高い医療の実現に向けた有識者検討会」が報告書を公表し、また、同年7月、厚生労働省・社会保険診療報酬支払基金（以下「支払基金」という。）が「支払基金業務効率化・高度化計画」をとりまとめた。これらの文書においては、ICTやAIの活用などにより、審査プロセスの見直しを含め、審査支払業務の効率化・高度化を目指すこととされ、国民健康保険中央会及び都道府県国民健康保険団体連合会（以下「国保連」という。）においても、支払基金における改革と同時並行的に、これと整合的かつ連携して取組みを進めることとされている。

この背景には、今後、高齢化が進むにつれ、後期高齢者医療のレセプト件数がますます増加していくことが予想される中、審査基準の統一化による質の維持された医療の確保、審査支払手数料の削減やレセプト審査の充実による国民負担の軽減、審査支払期間の短縮による医療機関への支払いの早期化など、国保連のレセプト審査業務の現状に対して関係者から出されている様々な要請がある。

国保連における現行のレセプト審査業務では、まず、ASPによる記載の不備等のチェックを行ったうえで、診療報酬点数表の基準を反映したルールベースのコンピュータチェックを行っており、全ての電子レセプトをチェックし、電子付せんを貼付することにより、基準を超えた診療行為が含まれるものの抽出を行っている。抽出されたレセプトについては、職員が目視点検を実施し、電子付せんを解除したり、医学的判断が求められるものについては疑義付せんを貼付し、審査委員に審査を依頼したり、また、審査委員会からの付託により査定又は返戻の処理を行っている。コンピュータチェックで電子付せんが貼付されなかったものであっても、職員が独自に問題点を洗い出すこともある。さらに、その後、医療の専門家からなる審査委員会で審査され、査定又は請求どおりの支払いという処理がなされている。

このように、審査全体では、記載の不備のチェック、保険診療のルールに照らした形式的なチェック、医学的な専門的観点からのチェックが総合的に行われ、査定又は査定対象外が判断されている。コンピュータチェックだけで審査は終了せず、多数の職員と審査委員が多大な時間と労力をかけて審査を行わなければならない状況が見られる。

現在、レセプトの電子化率がほぼ100%になっていることから、新たな技術を活用することにより、従来とは異なる審査手法の開発や新たな審査プロセスを構築できる可能性が生じている。例えば、今回の分担研究1. において、AIの技術を用いて生成を試みた患者像は、医療機関の属性を排した、それぞれの傷病に特徴的な診療行為の実施内容を表したものである。レセプト審査にこれを活用することにより、通常、実

施されることが少ない特異的な診療行為や処方を見ることができる可能性がある。

従来、審査効率化の手段としては、コンピュータチェックの拡充が主として議論されてきたが、この手法では大幅な業務の効率化を行うことは難しい。レセプト審査は、診療行為が保険診療ルール（療養担当規則、診療報酬点数表等）に適合するかどうかを確認する行為である。多様な患者に最適な医療を提供するという保険診療の性格上、保険診療ルールでは、医師等に一定の裁量を認めており、医師等の行為が保険診療ルールに適合するかどうかをコンピュータ等により機械的に判断することは不可能である。このため、専門的観点からの医学的妥当性の判断が必要となるが、AIに医学的判断を含む過去のレセプト審査結果のデータを学習させ、それを審査に活用することができれば、現在、職員や審査委員会が行っている業務の一部を代替し、業務の効率化を支援できる可能性がある。

このため、本分担研究では、レセプト審査業務の効率化・高度化を図る観点から、AIの技術を活用して審査を行う方法、それを実施するために解決されるべき課題等について検討を行う。具体的には、次の2つの活用方法について検証する。

- ① AIの技術を用いて生成された患者像の活用
- ② 審査済レセプトを学習させたAIの活用

本分担研究では、まず、B. において、研究方法を説明する。研究に当たっては、複数の国保連から匿名の審査済レセプトデータを借用し、様々な検証を行う。次に、Cにおいて、その結果を整理したうえで、Dにおいて考察を試みる。考察においては、現行のレセプト審査の業務フローを念頭に置きつつ、上記の2つの方法が現行の審査業務においてどのように活用できるのか具体的な検討を行う。また、AIの活用可能性のみならず、活用にあたって検討を深めるべき具体的な課題を提示する。

B. 研究方法

B.1 概要

A. で述べたAIの2つの活用方法の有効性について検証を行うため、次のような方法で研究を行った。

- ① 患者像からかい離したレセプト及びかい離したレセプトが多い医療機関の抽出

分担研究1. において生成した患者像を構成する項目は、患者カテゴリごとの傷病名、これと関連性のある摘要項目、回数数量等の平均的診療行為である。患者像に対応する平均的診療行為であれば、基本的には、保険診療ルールの範囲内であると考えられる。しかし、平均的診療行為の数値と比較して高い値を示す摘要項目のあるレセプトについては、患者像から外れた診療を行っている可能性があり、査定の可

能性があるレセプトとして、他に優先して審査する必要があるのではないかと考えられる。

このため、患者像からかい離したレセプトの抽出を試み、その内容を確認するとともに、かい離したレセプトが多い医療機関の抽出を行い、そのレセプトの内容を分析する。

② 審査済レセプトを学習させたAIの活用

患者像は使用せず、過去の審査済みのデータを機械学習させ、AIの判断を審査に活用できるかどうか、検証を行う。

B.2 定義

以下の分析に当たって使用する用語の定義は、次のとおりである。

- かい離度: 各レセプトに記録された摘要項目の数量や回数等ごとに、対応する患者像からの距離をそれぞれ算出。それらをレセプトごとに集約し、0~1の値として求めたもの
- かい離度上位のレセプト: かい離度が上位10%のレセプト
- かい離率: かい離度上位のレセプト数 / 総レセプト数
 - ※ 数のみだと大規模医療機関に偏るため、総レセプト数に対する割合を算出
- かい離度平均: 医療機関ごとに、かい離度上位のレセプトの、かい離度の平均を出したもの
 - ※ かい離率のみの場合、レセプトごとのかい離具合が反映されないため、かい離度も考慮
- 医療機関スコア: かい離度 * かい離度平均 * 10
 - ※ かい離率とかい離度平均の両方を考慮
- 査定レセプト数: 医療機関ごとの査定されたレセプト数
- 総レセプト数: 医療機関ごとのレセプト数
- 査定率: 査定レセプト数 / 総レセプト数

B.3 研究方法の詳細

①-1 患者像からかい離したレセプトの抽出

患者像からかい離したレセプトとして、生活習慣病の患者像の摘要項目における回数、数量に関してかい離したレセプトを抽出する。データは、1県の1か月分の医科、調剤レセプトを使用する。患者像からどれくらい離れているかを示すかい離度については、「Isolation Forest」という方法を使って算定する(枠内を参照)。

次に、以下の分析を行う。

- レセプト全体及びかい離度上位のレセプト(上位10%及び5%)の査定率の比較
- かい離度の階層ごとのレセプト数と査定率の関係

c. かい離度の最も高いレセプトの医学的見地からの確認

(参考) Isolation Forest について

患者像の標準から高い偏差を有するレセプトを同定するための教師なしアプローチ(予測の対象となる正解(教師)がない機械学習)である。各点が一つになるまで分割を行い、点を特定するまでの分割数が距離となる。点に到達するパスが短い(=早い段階で点が一つになる)ほど、外れ値になる可能性が高くなる。特徴空間におけるランダム分割をすることにより、点を分離するために決定木の最小経路長を決定することができる。この処理を何度も繰り返し、各点の経路長の期待値を設定した。算出式は、次のとおりである。

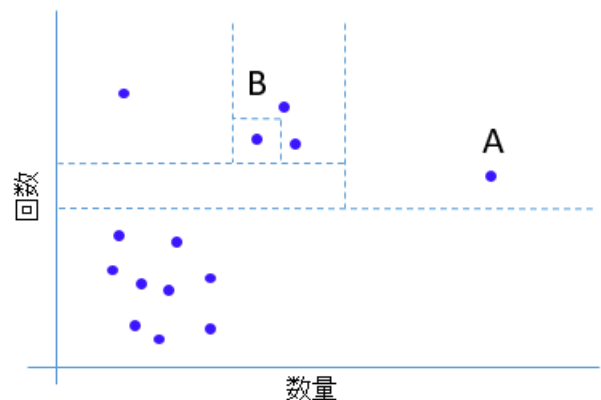
$$s(x, n) = 2 \frac{-E(h(x))}{c(n)}$$

$h(x)$ = 観測した経路長

$c(n)$ = n 個のノードを持つ失敗した二分探索木の平均経路長

かい離度算出のイメージは、図表21のとおりである。この中では、A点が最も早い段階で特定されるため、Aが最もかい離していると言える。

図表1 かい離度算出のイメージ



①-2 患者像からかい離したレセプトが多い医療機関の抽出

かい離したレセプトが多い医療機関の抽出を行い、以下の分析を行う。

- かい離度上位のレセプト数と査定レセプト数の関係
- かい離率と査定率の関係
- かい離率と総レセプト数の関係
- かい離度平均の階層ごとの医療機関の分布
- 医療機関スコアの階層ごとの医療機関の分布
- 査定率と医療機関スコアの関係

② 審査済レセプトを学習させたAIの活用

審査済みのレセプトデータを学習用データと判定用データに分ける。学習用データは、審査前の摘要データ及び補正・査定データとし、これらをAIに投入することにより、査定対象と査定対象外のデータの特徴を学習させる。

学習用データ以外のデータは判定用データとし、審査前の摘要データを学習後のAIに投入し、査定対象となるべきレセプトを判定させる。

AIによる判定後、判定データの補正・査定情報、査定後の摘要データを使用して、AIの判定結果が正しかったかどうかを確認する。この検証作業のイメージは、図表2のとおりである。

a. 使用したデータ

分担研究1. と同様、国民健康保険保険者及び後期高齢者医療広域連合の協力を得て、国保と後期高齢者医療のレセプトデータを使用する。

(種類)

医科、調剤のデータ

(データ期間、内容、件数)

分担研究1. と同じ。

(データ内容)

傷病名データ、査定前後の摘要データ、付せんデータ、補正・査定データ

b. 使用した技術

複数の機械学習アルゴリズム（ One-R, Ripper, Bayesian ）を組み合わせ、学習、判定を行う。

c. 検証方法

分析の妥当性の検証は、5-fold クロスバリデーション（5分割交差検証）を行い、結果を確認した。具体的には、使用するデータを5分割し、そのうちの1つをテスト事例とし、残る4つを学習事例とする。これを全事例が1回ずつテスト事例となるよう検証を繰り返し、5回の結果を平均して1つの推定を得る手法である。

d. 検証テーマと学習量

検証するテーマを2つ設定し、それぞれについて異なる学習データ量で判定を行い、結果を確認する（図表3）。

<②-1 全てのレセプトの学習>

AIに全てのレセプトを学習させ、査定対象と査定対象外の判定が可能か検証を行う。

使用するデータ量については、1 か月分、2か月分、4か月分とする。1 か月分のデータを使用する場合、学習データ7、判定データ3の割合で分割する。分割に

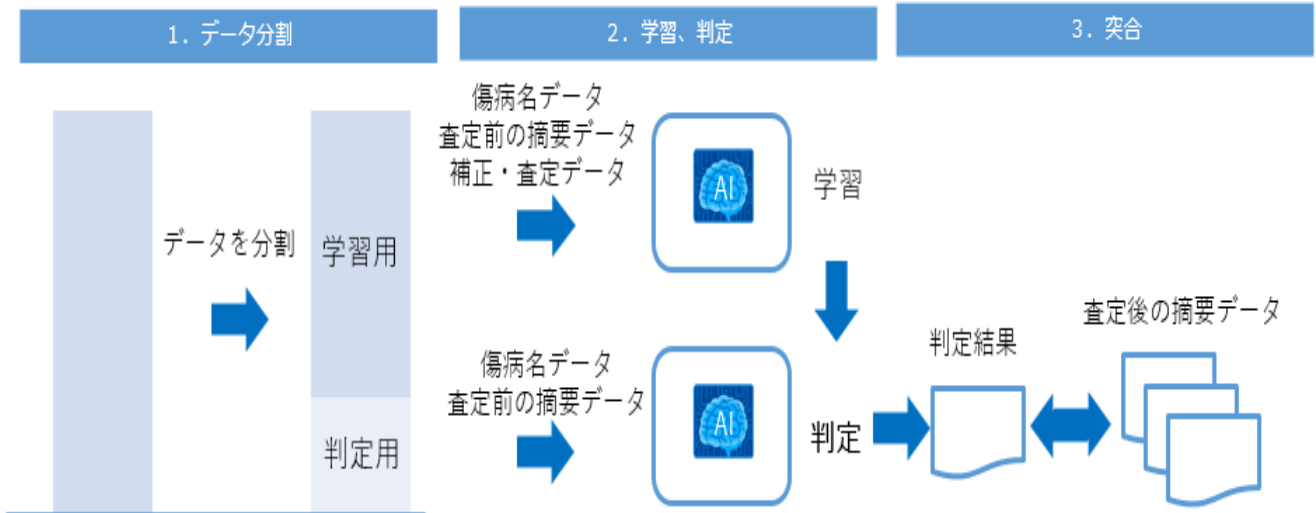
当たっては、査定されたレセプトが偏らないよう、ランダムに分割する。2か月分、4か月分のデータを使用する場合、診療月により分割する。学習データの件数については、図表4に示す。

<②-2 疑義付せんが貼付されたレセプトの学習>

AIに学習させる対象レセプトを疑義付せんが貼付されたものに限定する。現行の審査プロセスでは、コンピュータチェックによる電子付せんが貼付された後、職員が目視でその内容を点検し、医学的な判断を要する部分について疑義付せんを貼付し、審査委員に審査を依頼するが、その判定の結果、査定となる場合とならない場合がある。疑義付せんをどこに貼付するかは人が判断する部分であり、この内容をAIが学習し、査定となる箇所を判定できるか検証を行う。

使用するデータ量については、②-1と同じデータの2か月分と4か月分を診療月により分割して使用する。

図表 2 検証作業のイメージ



図表 3 レセプトの学習データと判定データへの分割

No.	検証テーマ	学習データ	判定データ	備考
② - 1	全てのレセプトの学習	0.7か月分	0.3か月分	5月分で学習、判定
		1か月分	1か月分	8月分を学習、9月分を判定
		3か月分	1か月分	6,7,8月分を学習、9月分を判定
② - 2	疑義付せんが貼付されたレセプトの学習	1か月分	1か月分	8月分を学習、9月分を判定
		3か月分	1か月分	6,7,8月分を学習、9月分を判定

図表 4 学習データと判定データの量 (1、2、4か月の場合)

1か月の場合		2か月の場合		4か月の場合	
A連合会 5月診療分	学習用 (0.7か月分)	A連合会 8,9月診療分	学習用 (8月診療分)	A連合会 6-9月診療分	学習用 (6-8月診療分)
1,778,874	1,778,874	3,570,769	1,801,105	7,187,806	5,418,142
	判定用 (0.3か月分)		判定用 (9月診療分)		判定用 (9月診療分)
	入院 16,585		入院 55,880		入院 55,880
	外来 516,362		外来 1,713,779		外来 1,713,779

C. 検証の結果

C.1 ①-1 患者像からかい離したレセプトの抽出

a. レセプト全体及びかい離度上位のレセプト（上位10%及び5%）の査定率の比較

図表5は、レセプト全体、かい離度上位10%、かい離度上位5%のレセプトについて、査定の有無を調べたものである。レセプト全体では1.02%が査定されていたが、かい離度上位5%では2.24%が査定されており、かい離度が高い層では、査定率が高いことがわかる。

b. かい離度の階層ごとのレセプト数と査定率の関係

図表6の棒グラフは、かい離度の階層ごとのレセプト数の分布を示したものである。また、折れ線グラフは、かい離度の階層ごとの査定率（査定されたレセプト数をその階層に属する総レセプト数で除したものを）を示している。かい離度が大きい階層ではレセプト数は少なくなるが、査定率は徐々に高くなり、かい離度1では査定率は5.23%に跳ね上がる。

c. かい離度の最も高いレセプトの医学的見地からの確認

かい離度1のレセプト172枚の中から6つのケースを無作為に抽出した（図表7）。6つのケースはもともと査定対象外であったが、分担研究1.の分担研究者が改めて医学的見地から確認を行った。その結果、一部レセプトで過剰と見受けられる診療行

為が見られたが、生活習慣病以外の傷病名があることなどから、査定されるべきであったと言えるようなものはなかった。したがって、かい離度が高いからと言って、直ちに査定対象になるわけではない。

C.2 ①-2 患者像からかい離したレセプトが多い医療機関の抽出

a. かい離度上位のレセプト数と査定レセプト数の関係

図表8は、医療機関ごとのかい離度上位のレセプト数と査定レセプト数をプロットしたものである。両者の間には正の相関関係が見られ（相関係数0.743）、かい離度上位のレセプト数の多い医療機関では、査定レセプト数も多いことがわかる。ただし、かい離度上位のレセプト数は大きいものの、査定レセプト数が小さい医療機関も見られた。

b. かい離率と査定率の関係

図表9は、かい離率と査定率の関係を示したものである。両者の間に相関関係は見られない。一部、かい離率は高いが、査定率が他よりも低い医療機関が見られた。

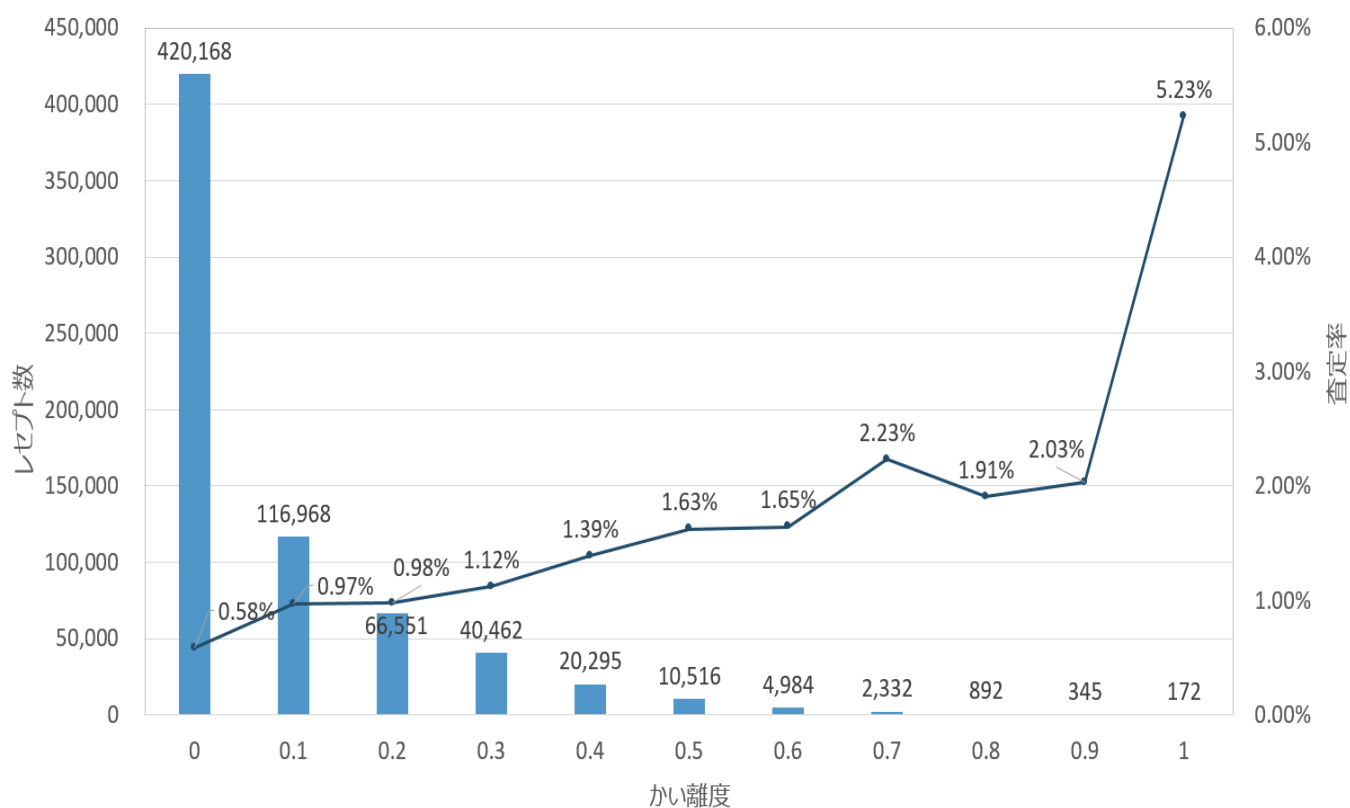
c. かい離率と総レセプト数の関係

図表10は、かい離率と総レセプト数の関係を見たものである。両者の間に相関関係は見られない。総レセプト数の少ない医療機関の中で、一部、かい離率が高いものが見られる。小規模な医療機関では、診療内容のばらつきが大きい可能性がある。

図表5 レセプト全体及びかい離度上位のレセプトの査定の有無



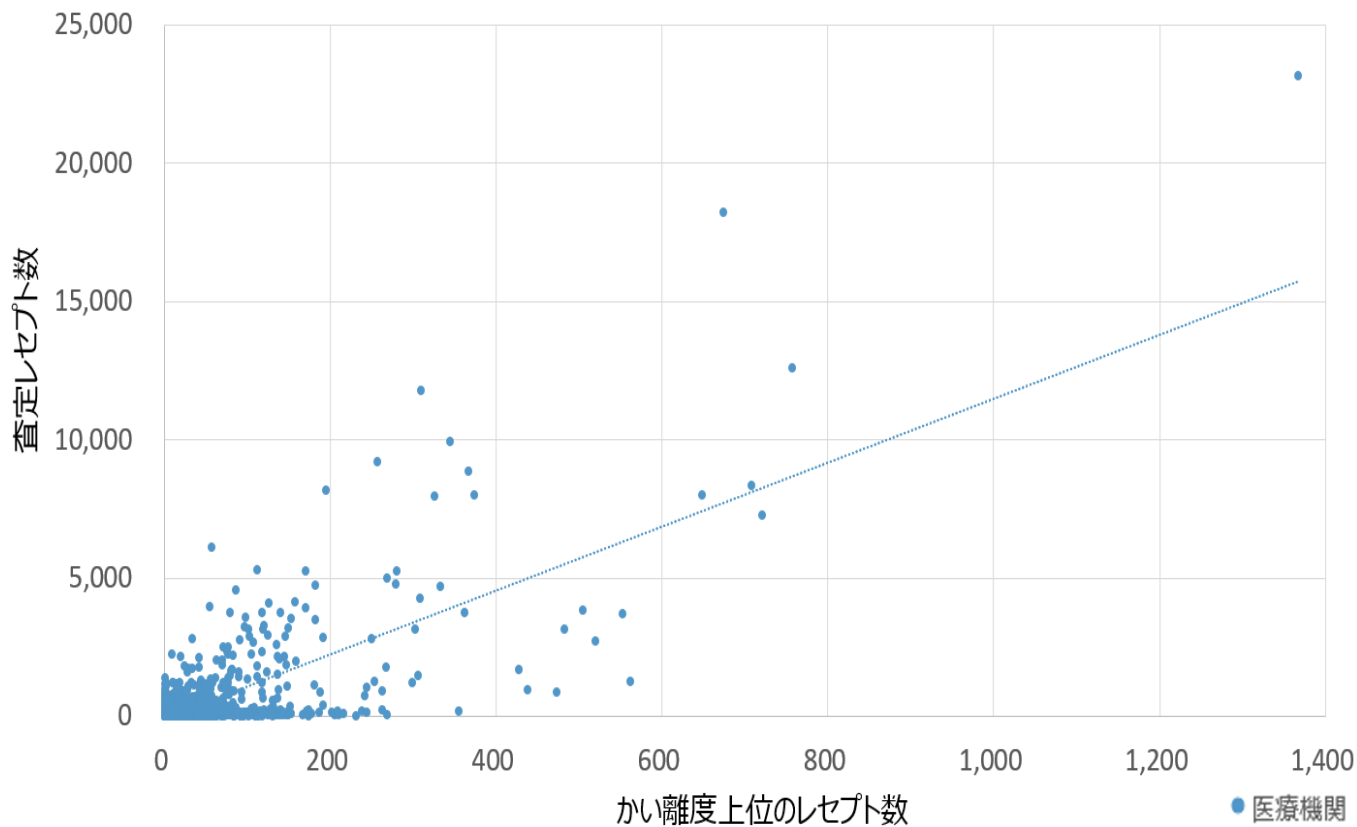
図表6 かい離度の階層ごとのレセプト数と査定率



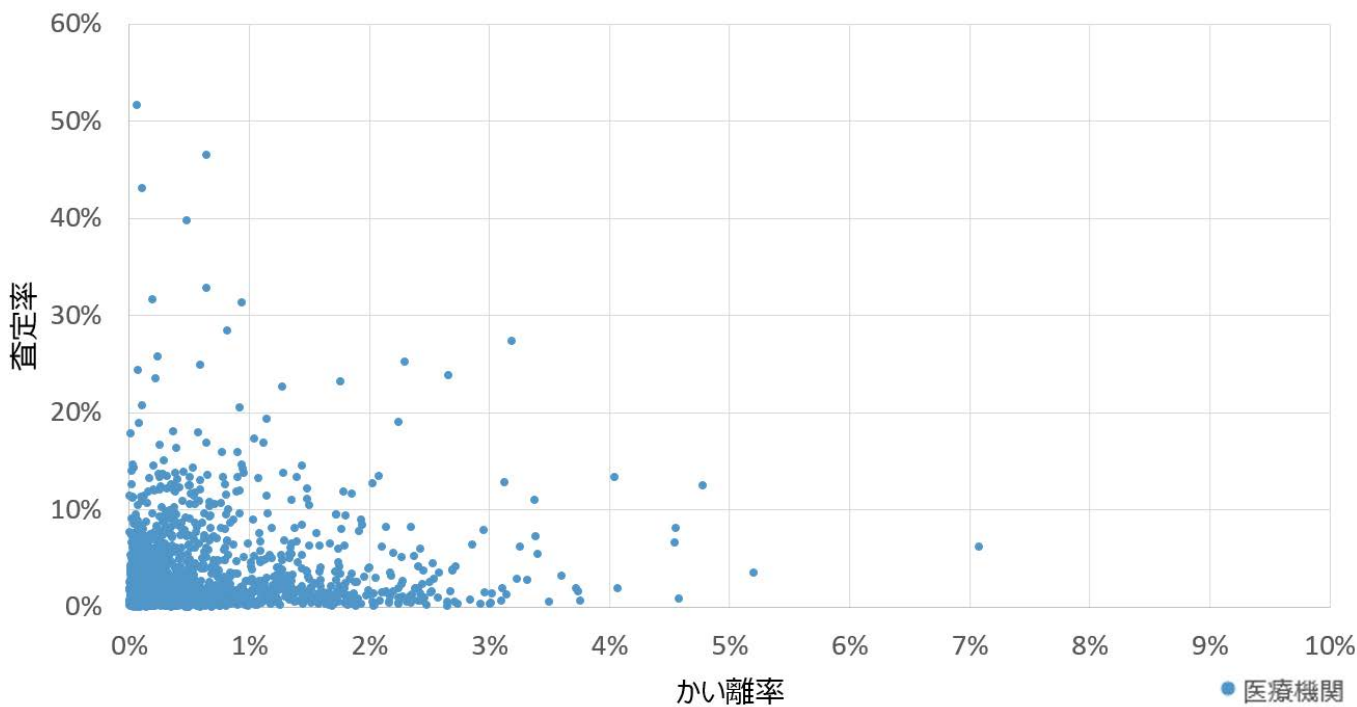
図表7 かい離度1のレセプトから抽出した事例

#	傷病名	年齢階層	初再診	入外	性別	医療資源投入量	かい離度	査定有無
1	糖尿病	75-89	再診	外来	男	低	1.0	査定なし
2	糖尿病	75-89	再診	入院	男	中	1.0	査定なし
3	高血圧	65-74	初診	外来	女	大	1.0	査定なし
4	高血圧	75-89	再診	入院	女	中	1.0	査定なし
5	脂質異常症	75-89	初診	外来	女	大	1.0	査定なし
6	脂質異常症	75-89	再診	入院	女	中	1.0	査定なし

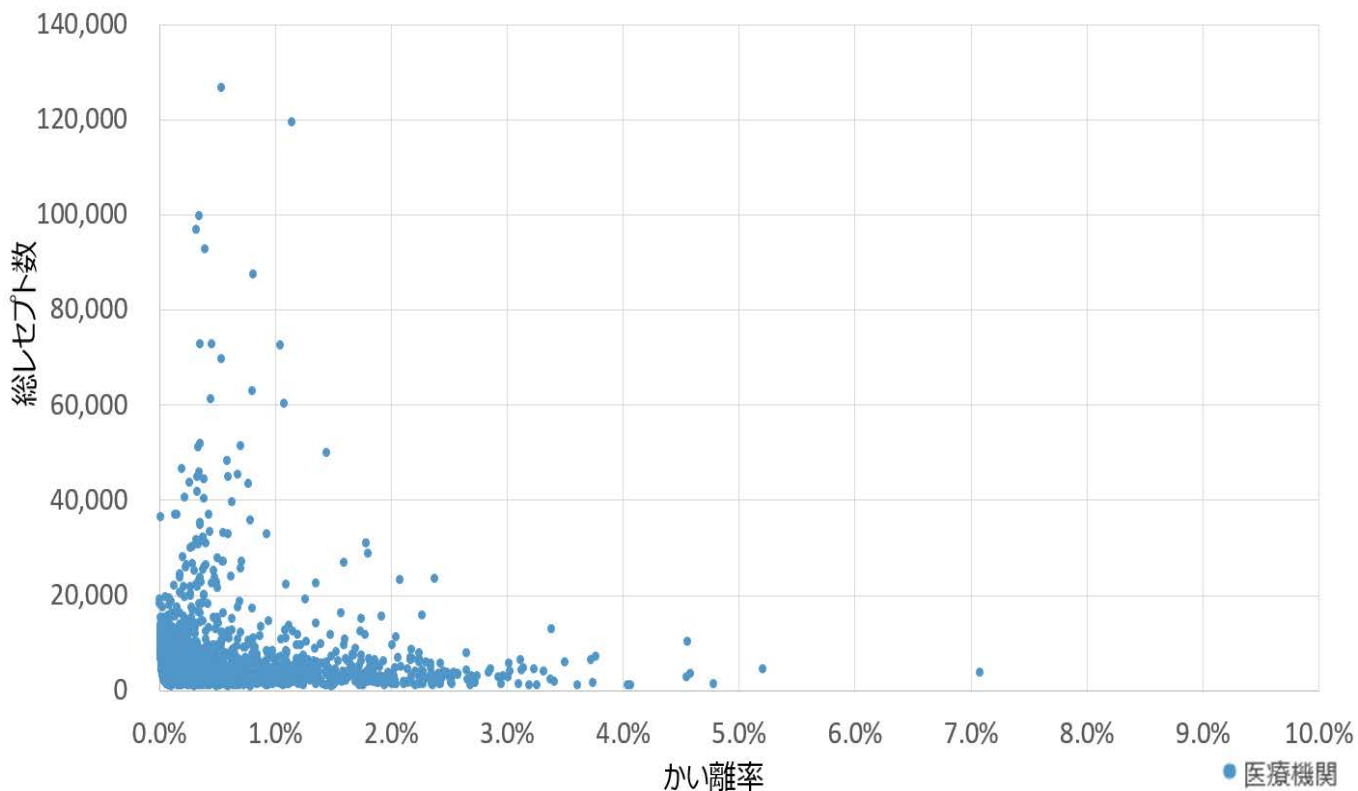
図表8 医療機関ごとのかい離度上位のレセプト数と査定レセプト数の関係



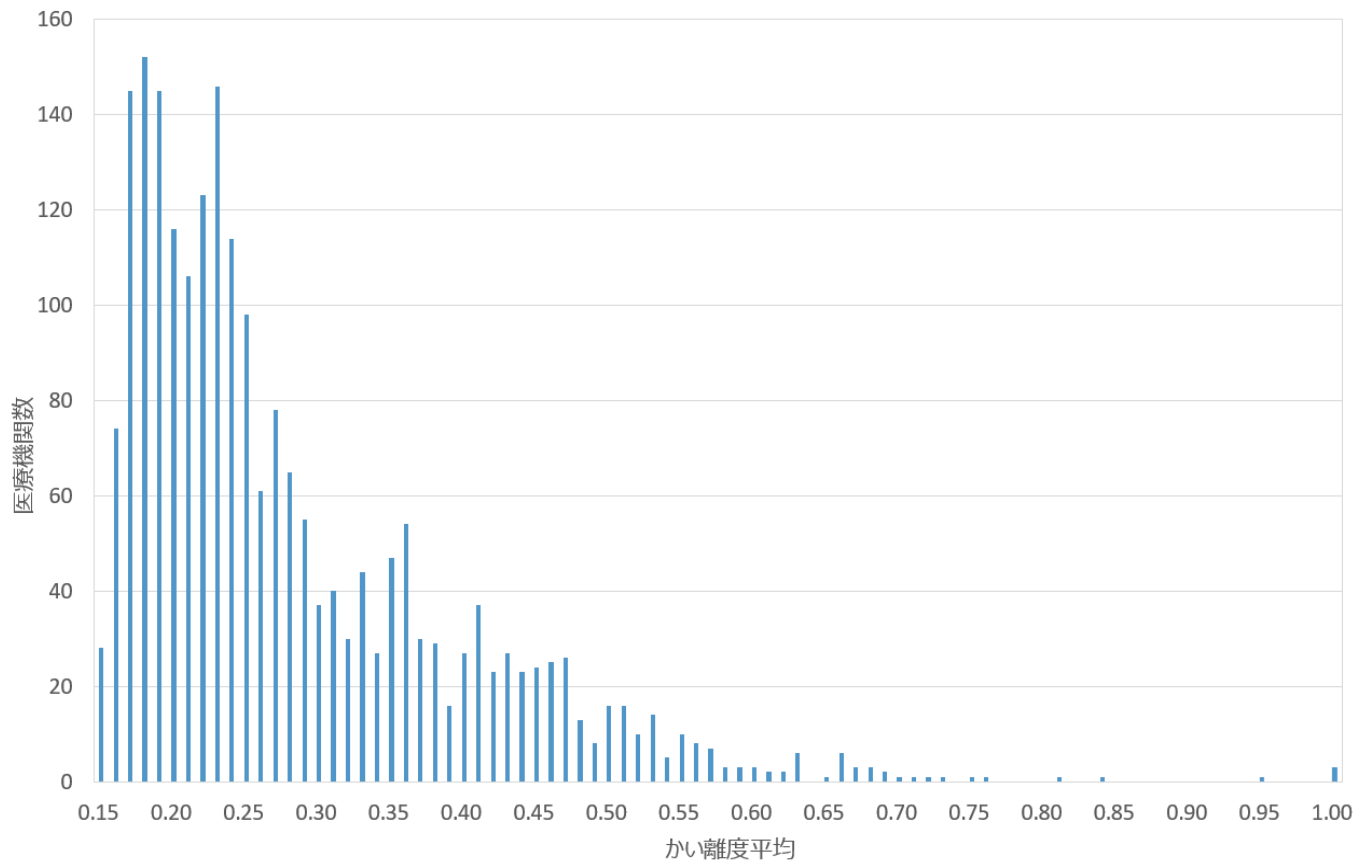
図表9 かい離率と査定率の関係



図表10 かい離率と総レセプト数の関係



図表11 かい離度平均の階層ごとの医療機関数



d. かい離度平均の階層ごとの医療機関の分布

図表11は、かい離度平均の階層ごとの医療機関数を示したものである。医療機関の多くは、かい離度平均0.1~0.3の間に含まれるが、一部、かい離度平均が大きい医療機関が見られる。

e. 医療機関スコアの階層ごとの医療機関の分布

図表12は、医療機関スコアの階層ごとの医療機関数を見たものである。医療機関のほとんどは、医療機関スコア0.02以下に含まれるが、一部、医療機関スコアが大きいものが見られる。

f. 査定率と医療機関スコアの関係

図表13は、査定率と医療機関スコアの関係を示したものであるが、相関関係は見られない。一部、医

(学習 1 か月分の場合)

入院の場合、実際に疑義付せんが貼付され、査定対象となったレセプトについての正解率は77.06%、実際には査定対象外であったレセプトの正解率は99.88%であった。外来では、それぞれ77.99%、99.96%であった(図表17)。

(学習 3 か月分の場合)

入院の場合、実際には疑義付せんが貼付され、査定対象となったレセプトの正解率は78.03%、実際には査定対象外であったレセプトの正解率は99.88%であった。外来では、それぞれ78.49%、99.97%であった(図表18)。
療機関スコアが大きい、査定率が低い医療機関が見られる。

C.3 ②機械学習による査定対象レセプトの判定

<②-1 全てのレセプトの学習>

図表14~16は、実際には査定対象(外)であったレセプトについて、AIが査定対象(外)と正しく判定した割合(正解率)を示したものである。

(学習 0.7 か月分の場合)

入院の場合、実際には査定対象であったレセプトについての正解率は83.20%、実際には査定対象外であったレセプトについての正解率は98.88%であった。外来の正解率については、それぞれ86.28%、99.87%であった(図表14)。

(学習 1 か月分の場合)

入院の場合、実際には査定対象であったレセプトについての正解率は87.79%、実際には査定対象外であったレセプトの正解率は98.96%であった。外来については、それぞれ92.57%、99.88%であった(図表15)。

(学習 3 か月分の場合)

入院の場合、実際には査定対象であったレセプトについての正解率は87.94%、実際には査定対象外であったレセプトについての正解率は99.01%であった。外来では、それぞれ92.60%、99.93%であった(図表16)。

<②-2 疑義付せんが貼付されたレセプトの学習>

図表17・18は、疑義付せんが貼付され、実際に査定対象(外)となったレセプトについて、AIが査定対象(外)と正しく判定した割合(正解率)を示したものである。

<学習量と判定精度の関係>

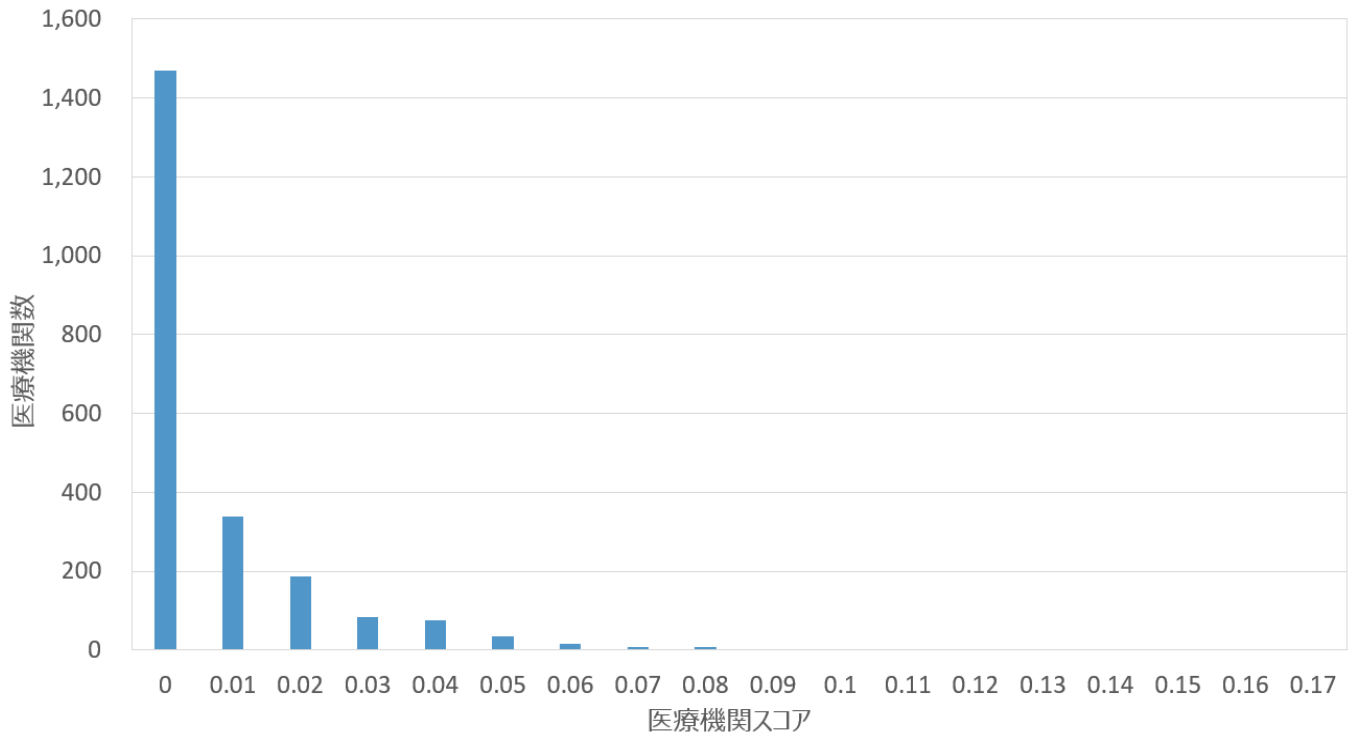
図表19・20は、機械学習の量とAIによる判定の精度の関係を示したものである。

全てのレセプトを学習させた場合、査定対象(査定対象外)と判定したが、実際は査定対象外(査定対象)であったものの不正解の件数は、AIに1か月分を学習させた場合よりも、3か月分を学習させた方が少なくなった(図19)。これは、疑義付せんが貼付されたレセプトを学習させた場合も同様である。

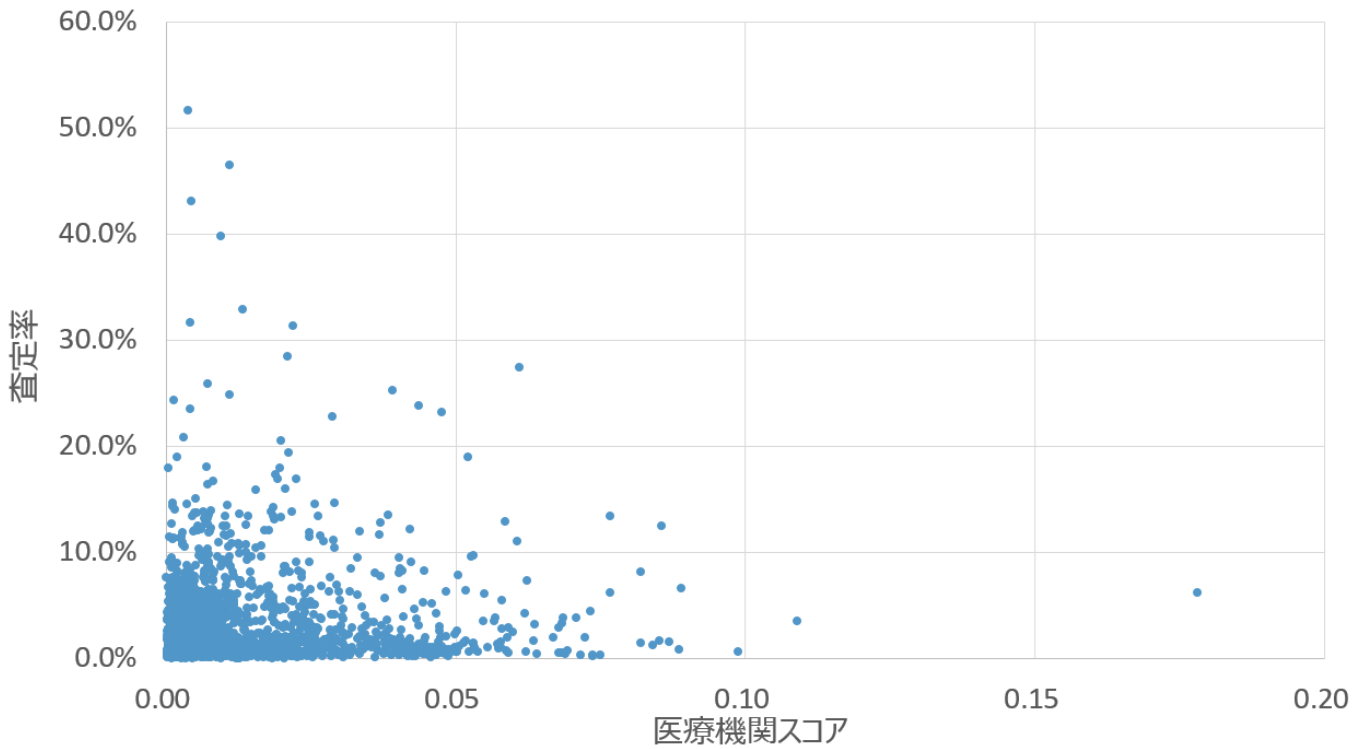
また、全てのレセプトを学習させた場合において、AIに学習させた審査済レセプトのうち実際には査定されていたものについての正解率を見ると、入院での学習0.7か月分の正解率は83.20%であるが、1か月分では87.79%と、4.59%と大きく増加した(図表20)。さらに、学習量を3か月分に増やすと、正解率は87.94%となったが、1か月分からの増加分は0.15%のみであった。学習を増やすと正解率は増加する傾向が見られるものの、その増加分は小さくなっている。外来の場合も同様に学習を増やすと、正解率は増加するが、その増加分はわずかである。これは、外来の正解率は0.7か月分であっても既に99.9%程度となっているためであり、学習を増加させることによる改善の余地は小さい。

疑義付せんが貼付されたレセプトについても、1か月分から3か月分に学習を増やせば、若干ではあるが、正解率が上昇する傾向が見られた。

図表12 医療機関スコアの階層ごとの医療機関数



図表13 査定率と医療機関スコアの関係



図表14 AIによる判定結果（全てのレセプトの学習 - 学習 0.7か月分）

入院 / 外来	実際の審査結果 (計 16,752 件)		AI による判定結果		正解率
			査定対象外	査定対象	
入院	査定対象外	14,936件	14,769 件 (正解)	167 件 (不正解)	98.88 %
	査定対象	1,816件	305 件 (不正解)	1,511 件 (正解)	83.20 %

入院 / 外来	実際の審査結果 (計 517,017 件)		AI による判定結果		正解率
			査定対象外	査定対象	
外来	査定対象外	506,859件	506,204 件 (正解)	655 件 (不正解)	99.87 %
	査定対象	10,158件	1,396 件 (不正解)	8,762 件 (正解)	86.26 %

図表15 AI による判定結果（全てのレセプトの学習 - 学習 1か月分）

入院 / 外来	実際の審査結果 (計 55,880 件)		AI による判定結果		正解率
			査定対象外	査定対象	
入院	査定対象外	49,786件	49,266 件 (正解)	520 件 (不正解)	98.96 %
	査定対象	6,094件	744 件 (不正解)	5,350 件 (正解)	87.79 %

入院 / 外来	実際の審査結果 (計 1,713,779 件)		AI による判定結果		正解率
			査定対象外	査定対象	
外来	査定対象外	1,671,634件	1,669,696 件 (正解)	1,938 件 (不正解)	99.88 %
	査定対象	42,145件	3,131 件 (不正解)	38,277 件 (正解)	92.57 %

図表16 AI による判定結果（全てのレセプトの学習 - 学習 3か月分）

入院 / 外来	実際の審査結果 (計 55,880 件)		AI による判定結果		正解率
			査定対象外	査定対象	
入院	査定対象外	49,786件	49,295 件 (正解)	491 件 (不正解)	99.01 %
	査定対象	6,094件	735 件 (不正解)	5,359 件 (正解)	87.94 %

入院 / 外来	実際の審査結果 (計 1,713,779 件)		AI による判定結果		正解率
			査定対象外	査定対象	
外来	査定対象外	1,671,634件	1,670,445 件 (正解)	1,189 件 (不正解)	99.93 %
	査定対象	42,145件	3,118 件 (不正解)	39,027 件 (正解)	92.60 %

図表17 AI による判定結果（疑義付せんが貼付されたレセプトの学習 - 学習 1か月分）

入院 / 外来	実際の審査結果 (計 55,880 件)		AI による判定結果		正解率
			疑義&査定対象	疑義&査定対象外	
入院	疑義&査定対象	510件	393 件 (正解)	117 件 (不正解)	77.06 %
	疑義&査定対象外	55,370件	67 件 (不正解)	55,303 件 (正解)	99.88 %

入院 / 外来	実際の審査結果 (計 1,713,779 件)		AI による判定結果		正解率
			疑義&査定対象	疑義&査定対象外	
外来	疑義&査定対象	5,985件	4,668 件 (正解)	1,317 件 (不正解)	77.99 %
	疑義&査定対象外	1,707,794件	708 件 (不正解)	1,707,086 件 (正解)	99.96 %

図表18 AI による判定結果（疑義付せんが貼付されたレセプトの学習 - 学習 3か月分）

入院 / 外来	実際の審査結果 (計 55,880 件)		AI による判定結果		正解率
			疑義&査定対象	疑義&査定対象外	
入院	疑義&査定対象	510件	398 件 (正解)	112 件 (不正解)	78.03 %
	疑義&査定対象外	55,370件	66 件 (不正解)	55,304 件 (正解)	99.88 %

入院 / 外来	実際の審査結果 (計 1,713,779 件)		AI による判定結果		正解率
			疑義&査定対象	疑義&査定対象外	
外来	疑義&査定対象	5,985件	4,698 件 (正解)	1,287 件 (不正解)	78.49 %
	疑義&査定対象外	1,707,794件	563 件 (不正解)	1,707,231 件 (正解)	99.97 %

図表19 AIの判定が不正解となったレセプト件数（学習 1か月分、3か月分）

検証テーマ名	不正解の内容	入院 / 外来	学習 1か月分	学習 3か月分	増減
全てのレセプト の学習	査定対象外と判定したが、 実際は査定対象	入院	744	735	-9
		外来	3,131	3,118	-13
	査定対象と判定したが、 実際は査定対象外	入院	491	491	0
		外来	1,938	1,189	-749

検証テーマ名	不正解の内容	入院 / 外来	学習 1か月分	学習 3か月分	増減
疑義付せんが 貼付された レセプトの学習	査定対象と判定したが、 実際は査定対象外	入院	67	66	-1
		外来	708	563	-145
	査定対象外と判定したが、 実際は査定対象	入院	117	112	-5
		外来	1,317	1,287	-30

図表20 入院と外来における正解率（学習 0.7か月分、1か月分、3か月分）

検証テーマ名	AIに学習させた審査済レセプト		正解率(%)		
			学習0.7か月分 の場合	学習1か月分 の場合	学習3か月分 の場合
全てのレセプト の学習	入院	実際には査定されていたレセプト	83.20	87.79	87.94
		実際には査定対象外であったレセプト	98.88	98.96	99.01
	外来	実際には査定されていたレセプト	86.26	92.57	92.60
		実際には査定対象外であったレセプト	99.87	99.88	99.93
疑義付せんが 貼付されたレ セプトの学習	入院	実際には査定されていたレセプト	—	77.06	78.03
		実際には査定対象外であったレセプト	—	99.88	99.88
	外来	実際には査定されていたレセプト	—	77.99	78.49
		実際には査定対象外であったレセプト	—	99.96	99.97

D. 考察

D.1 AIにより生成した患者像の審査への活用

(1) 患者像からかい離したレセプトの重点的審査

検証作業において、かい離度が最も高い（かい離度 1）レセプトの中から実際には査定されなかった事例を抽出し、医学的見地から改めて確認を行ったが、審査結果は妥当なものと考えられた。かい離度が高くても、保険診療ルール の範囲内にとどまっている限り、査定対象とはならない。このため、かい離度が高いことが直ちに査定に結び付くわけではない。

しかし、検証結果が示すとおり、かい離度が1のレセプトについては査定率が高い傾向が見られることから、レセプト審査に当たり、特に注意を払う必要があると考えられる。

なお、本研究で生成した患者像は、生活習慣病のみである。1枚のレセプトに生活習慣病のみが記録されるということはまれであり、通常、他の傷病名が含まれていることが多い。また、1つの診療行為が2つ以上の傷病に基づいて行われたことも考えられる。そのため、生活習慣病以外の患者像も生成した上で、かい離度を活用することの有効性を検討する必要があると考えられる。

(2) 患者像からかい離したレセプトが多い医療機関の重点的審査

検証作業において医療機関スコアを算出したところ、一部、医療機関スコアは大きい が、査定率が低い医療機関が見られた。これについても、かい離度の高い診療を行う傾向があったとしても、保険診療ルー

ルの範囲内である限りは査定対象とならないことから、医療機関スコアの大きさが直ちに査定に結び付くということではない。

しかし、医療機関スコアが大きく、かつ、かい離度1のレセプトが多い医療機関を重点的に審査することにより、審査の効果を向上させる可能性がある。ただし、前述のとおり、生活習慣病以外の患者像を作成した上で、かい離度等の判定の精度を高めることが必要である。

D.2 AIの判定精度の向上

(1) 学習量の増加

検証の結果、AIによる判定について、高い精度で行われうる可能性が確認されたが、それでもその判断が不正解となる場合がある。AIに実際には査定対象外であったものを判定させた場合、100%に近い正解率が得られたが、実際には査定対象だったレセプトの場合、正解率は90%前後にとどまる。機械学習の性質上、AIの不正解をゼロにすることは困難であるが、できるだけ減らすことが望ましい。

検証作業の結果、AIの正解率は、学習量が多いほど高いことがわかった。学習量のある程度増やすことにより、判定精度を高めることができると考えられる。ただし、一定以上、学習量を増やしたとしても、判定精度向上の限界効果は小さくなる可能性がある。いずれにしても、今後、学習量を更に増やした検証を行い、判定精度の改善度合いを確認する必要がある。

なお、レセプトは、インフルエンザや花粉症等、季節性を持つほか、国保連において、審査の観点や手法を毎月見直ししているため、学習データの傾向が毎月変動する要素が存在する。

また、定期的な診療報酬改定や新薬の使用開始に

よる算定ルールの変更により、学習すべき内容を変更する必要が生じると考えられる。

このため、これらの要素を考慮し、また、これらを補う他の審査手法との組み合わせも検討しながら、最適な学習量を見極める必要があると考えられる。

(2) 学習・判定の精緻化

本研究では、レセプト単位での判定について検証を行ったが、学習項目を追加し、判定プロセスに処理を加えることにより、摘要コード単位で査定事由を判定することが可能となるほか、査定の内容をAIの判断に委ねることも可能になると考えられる。つまり、レセプトが査定されたかどうかに加え、それらがどのような値に減点されたかも学習させることで、摘要項目の数量、回数の査定内容を判断することができると考えられる。

(3) 再審査結果の学習と一次審査(原審査)への反映

国保連での審査を終えたレセプトは、請求支払計算が行われた後、各保険者に送られる。各保険者はレセプト点検を実施しており、国保連の審査結果に疑義がある場合、必要に応じて再審査請求を行っている。また、保険医療機関が再審査請求を行うこともある。再審査請求を受領した国保連では、再審査を行い、容認又は原審どおりの判定を行っており、容認した場合には、その内容を一次審査(原審査)の審査観点に反映させている。

今回の検証では、AIには一次審査(原審査)の結果のみを学習させたが、再審査の結果も学習させることにより、その視点もAIの判定に自動的に取り込むことができるため、判定精度を向上させることができると考えられる。

D.3 AIの活用範囲

本研究では、AIに全レセプトを学習させたほか、疑義付せんが貼付されたレセプトを学習させ、AIの活用可能性を検証した。しかし、AIを活用するデータの範囲については、他にも様々なものが考えられる。全てのレセプトを学習させる方法だけでなく、例えば、判定精度が高い一定の分野、AIの活用により効率化効果が上がると考えられる分野などに絞って学習させ、判定を行わせる方法もありうる。

この場合、活用の目的により学習するデータの量や内容が異なることから、全てを単一のモデルで行うのではなく、それぞれの学習、判定に特化したモデルを構築し、組み合わせることで、判定の精度を改善することができると考えられる。

(査定対象外となるべきレセプト)

査定対象外となるべきレセプトについてのAIの正解率は100%に近い水準にあり、判定精度が高いことから、AIが査定対象外と判定したレセプトについて

は、職員の点検や審査委員による審査に回さないという取扱いも考えられる。査定対象外となるレセプトの全レセプトに占める割合は、入院では9割程度、外来では97~98%程度であり、これらの判定をAIに任せ、職員や審査委員は、査定対象と判定されたレセプトの点検・審査に専念することとすれば、業務の合理化を図ることができるのではないかと考えられる。

(外来レセプト、高額でないレセプト)

検証作業の結果、入院と外来におけるAIの正解率は、若干ではあるが、外来の方が高かった。これは、データ量の違い(外来のデータ件数が入院より圧倒的に多い)に加え、外来の場合、内容が単純なレセプトが多いためと考えられる(図表20)。

一般的には、レセプトの内容が複雑になると、AIの判定精度が低くなる可能性がある。このため、AIのレセプト審査への活用は、外来など内容が比較的単純なレセプトから実施していくことも考えられる。

請求点数が高いレセプトについても、内容が複雑性を増すことから、同じことが言える。このため、例えば、AIに学習、判定させるレセプトを、外来で、かつ、高額でないものに絞ることとすれば、AIの判定精度を高めることが期待できる。

(コンピュータチェックでは電子付せんが貼付されなかったレセプト)

現行の業務では、コンピュータチェックによる電子付せんが貼付されなかったレセプトについて、職員が独自に問題点を洗い出すとともに、審査委員が専門的見地から審査を行っている。

今回の検証では、AIに「職員が疑義付せんを貼付したレセプト」を学習させてみたが、コンピュータではなく、人が判断すべき部分に絞って学習、判定を行わせることで、業務の効率化を図ることができる可能性がある。

(特定の診療科のレセプト)

耳鼻科、眼科、産婦人科等、各国保連で対象レセプトを区分することが可能な単位でAIに学習、判定を行わせることにより、判定精度が高い診療科からAIを活用していくことも考えられる。

(DPC、歯科のレセプト)

今回の検証に当たっては、医科と調剤レセプトのみを使用した。AIをDPCや歯科の分野で活用することも考えられる。DPCの出来高部分については、医科の入院と同様、活用可能と考えられる。歯科の場合、医科と同様、コンピュータチェック以外に人間の判断が必要とされる部分に活用できる可能性がある。

(縦覧点検、横覧点検、突合点検)

今回の検証では、単月レセプトに焦点を当て、査定対象・対象外となった内容を学習させたが、現行の業

務では、同一被保険者のレセプトを連月や入院、外来、医科、調剤等で連結してコンピュータチェックを実施し、点検・審査を行っている。

これらのデータについても併せてAIに学習させることにより、判定精度を向上させることができると考えられる。

D.4 AIの審査業務への活用のあり方

現行の業務では、コンピュータチェックによる電子付せんが貼付されたレセプトを職員が点検し、その中で医学的判断が必要なレセプトは疑義付せんを貼付し、審査委員に審査を依頼する。審査委員会は、医学的見地からレセプト内容の妥当性を判断し、最終的に決定する。

今回の検証により、AIに過去の審査データ、すなわち、コンピュータチェックや職員・審査委員の判断を通じて行われた最終的な審査結果を学習させれば、新たなレセプトが査定対象となるかどうか、ある程度の精度で判定できることがわかった。

AIを有効に活用すれば、点検・審査に係る事務量を大幅に減らすことも可能になろう。これにより、職員の点検や審査委員による審査に必要な工数を削減し、より詳細な点検・審査が必要なレセプトに時間を割

く等、審査の効率化と充実を支援することができると考えられる。

また、現在の業務においては、医学的判断が必要な審査の一部について、審査委員から職員に妥当性の確認、査定内容の判断が付託される場合があるが、これらの部分だけでもAIに判定させることにすれば、職員の作業工数を軽減することができると考えられる。

なお、コンピュータチェックによる電子付せんの貼付は、AIによる判定とは機能が異なるため、AIを導入するとしても、引き続き実施することが必要と考えられる。コンピュータチェックは、診療報酬点数表等に定められた算定ルール等が守られているかどうかを形式的にチェックすることを目的としている。そのようなチェックだけであれば、AIも対応することができるが、算定ルールが変更されたり、新たな医薬品が保険適用されたりした場合などには、即座に対応することができない。審査済のレセプトが一定量蓄積されるまでは学習することができないためである。

現行のレセプト審査の業務フローを踏まえ、AIをどのように活用できるのか具体的なイメージを図表21に示す。

図表21 AIのレセプト審査業務への活用のイメージ

現行の業務フロー		AIの具体的な活用方法
3か月～半年毎 月次処理 随時	①医療機関の評価	重点的に審査すべき対象として、AIの技術を活用し、かい離度の高いレセプトの多い医療機関を抽出することができる。
	②受付・ASPによる点検	
	③コンピュータチェック	コンピュータチェックは、AIと併用する。
	④審査事務共助	AIを活用することによって効率化できる業務の例 ・AIが査定対象外と判定したレセプトの点検 ・電子付せん貼付レセプトへの疑義付せんの貼付 ・審査委員会から職員に付託された査定処理 ・電子付せんのないレセプトの問題点の洗い出し ・外来レセプトや高額でないレセプトの点検 ・特定の診療科のレセプトの点検
	⑤審査委員による審査	AIを活用することにより効率化できる業務の例 ・AIが査定対象外と判定したレセプトの審査 ・外来レセプト、高額でないレセプトの審査 ・特定の診療科のレセプトの審査
	⑥審査後処理	
	⑦再審査	AIの判定を活用する場合、AIの判定誤りの再審査を行う。 AIは再審査結果を学習し、一次審査の精度を高めることができる。

D.5 AI活用の費用対効果

AI をレセプト審査業務に導入する場合、その費用対効果を推計し、審査事務共助等の効率化、審査委員の拘束時間の短縮によって浮いた資源を、医学的判断への専念等、審査の充実に回すことによる医療費の削減などの効果が導入費用を上回ることが条件となる。

AIに実質的な判断を任せる場合、AI による判定の誤り（不正解）をゼロにすることは困難であるため、判定精度の高いレセプトのみに活用することとしたとしても、本来、査定対象のものを誤って査定対象外としてしまうことも起こりうる。導入効果の算定に当たっては、査定の精度を向上させる工夫を重ねる必要があるが、AIが判定を誤ることによるコストも考慮しておく必要がある。

D.6 AI活用による審査結果の説明責任

審査の結果、査定となった箇所を医療機関に知らせる際、その判断の根拠を示すことが求められるが、AIが往々にしてブラックボックス化し、なぜそのように判定されたのか理由がわからないといった状態になるおそれもある。

本研究では、レセプト単位で査定対象の抽出を行ったが、学習項目の追加や判定プロセスに処理を追加することにより、摘要コード単位で査定事由を判定することが可能になると考えられる。

現状では、職員や審査委員が査定事由コード（A. 不適応 B. 過剰 C. 重複 D. 不相当、不必要）を登録し、これを医療機関に通知する仕組みとなっているが、このコードをAIに学習させることにより、査定・査定対象外の判定のほか、当該摘要コードの査定事由を提示することができるようになる。

ただし、実質的にAIの判断に基づき査定を行うことになる場合、AIの判断に誤りが紛れ込む可能性があるため、再審査請求の体制を整えておくことが必要である。

D.7 審査における地域差の確認

患者像や機械学習を活用することで、審査における地域差の解消を支援できる可能性がある。

患者像は1都道府県の国保、後期データ全体を使用して作成されたものであり、その都道府県の平均的な診療内容を表していることから、これを活用して、国保連間の審査基準の差異（ローカルルール）の解消を推進することができると考えられる。

具体的には、疾病ごとに生成された患者像を都道府県間で比較することにより、地域差の存在が明確化され、地域差を踏まえた上で、審査基準の差を比較することで、審査における地域差の解消に向けた議論に資することができると考えられる。

更によれば、機械学習の活用において、47国保連のレセプトデータを学習させることができるのであれば、それが医学的に見て妥当かどうかは別として、形の上では全国的な統一基準を作成することも可能になる。

E. 結論

本分担研究では、レセプト審査業務におけるAIの活用の有効性について検証を行い、具体的な活用のあり方について検討を行った。

患者像の活用については、今回の研究では、生成された患者像が一部の疾病に限られていたことに加え、患者像からかい離したレセプトが直ちに査定対象になるわけではないことから、これを直接審査に活用することはできないものの、重点的に審査を行う対象として、かい離度の高いレセプトやそれらの数が多い医療機関を抽出することには役立つと考えられる。

また、AIに審査済みのレセプトを学習させ、査定対象（外）の判定を行わせることについては、更に判定精度を向上させるための努力が必要であるものの、一定の有効性が確認された。AIが活用できる分野や業務については、今回実施した検証テーマ以外にも様々なものが考えられることから少なくとも今回使用したデータと同規模のデータを使用し検証作業を続けていくことが必要である。

今後、高齢化の進行により国保連や国保中央会では、審査しなければならぬレセプトの一層の増加が予想されるが、国民負担を軽減する観点からは、審査業務の効率化・高度化は、待ったなしの状況にあると言える。また、審査委員は、医療機関における診療等の通常の業務を行いながら、夜間や休日に国保連や国保中央会において審査業務に従事しており、AIを活用し、審査の効率化を図ることにより、その負担を軽減することができると考えられる。医療機関にとっても、AIの活用により効率的な審査が可能になれば、迅速な支払いが可能になるといったメリットがある。

AIのレセプト審査業務への導入に当たっては、解決すべき課題はあるものの、活用できる範囲が大きくなれば、それに応じてメリットも大きくなると考えられることから、各国保連や国保中央会において、審査委員、保険者、医療機関など関係者の理解を得ながら、活用に向けた具体的な検討を加速することが望まれる。

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

名称：審査支援システム、審査支援方法及び
審査支援プログラム

種類：特許権

番号：特願2018-240591

出願年：2018

2. 実用新案登録

なし

3. その他

なし

