



Title: Care-needs level prediction for elderly long-term care using insurance claims data
(レセプトデータを活用した高齢者介護の要介護度予測)

Authors: Hiroaki Fukunishi, Yasuki Kobayashi
(福西広晃 (東京工科大 コンピュータサイエンス学部 講師)、小林康毅 (東京大学 大学院医学系研究科))

Journal: Informatics in Medicine Unlocked 41 (2023) 101321

掲載年月: 2023 年 8 月

研究概要: 本研究は、75 歳以上の高齢者の将来の要介護度を予測する機械学習モデルを開発することを目的とする。学習データとして、約 4 万 8 千人の高齢者の介護レセプトと医療レセプトを使用した。予測モデルは、516 個の特徴量 (年齢、性別、514 の疾患など) を説明変数とし、4 クラスに分類した介護度を予測することとした。機械学習アルゴリズムには、解釈可能性を持つ多項ロジスティック回帰アルゴリズム(MLR)を採用した。MLR の予測性能は、他の機械学習アルゴリズムに匹敵する高い性能を示した。この研究の成果は、公的な健康保険制度と介護保険制度の被保険者のデータに基づいて、要介護の高リスク地域を把握するために利用できる可能性がある。

研究背景: 日本では高齢化に伴い要介護認定を受ける高齢者の数が増加し、医療および看護ケアの需要、および介護給付の費用が着実に増加している。今や、高齢者の長期介護は日本における重要な社会課題となっている。日本政府は医療や介護のレセプトデータから要介護リスクの高い高齢者を分析し、介護予防施策の実施を推進している。近年、著しい進歩を遂げている機械学習技術を用いて要介護の高リスク者を予測できれば、介護予防施策を支援することが可能となる。

研究成果: 多項ロジスティック回帰アルゴリズム (MLR)の予測性能は、加重平均の適合率、再現率、F 値、およびリフトスコアでそれぞれ 0.694、0.505、0.567、および 1.333 を達成した。この結果は、他の機械学習アルゴリズムであるサポートベクトルマシン(SVM)およびランダムフォレスト(RF)に匹敵する性能であった。MLR が SVM と RF と大きく異なる点は、MLR の予測モデルが解釈性を持つことである。MLR のモデル係数を解釈することにより、各介護度に影響を与える要因を分析することに成功した。各介護度の上位 3 個の要因は以下の通りとなった：低介護度クラス：女性、高血圧、膝蓋骨症状；中介護度クラス：年齢、アルツハイマー型認知症、膀胱の神経筋機能障害；高介護度クラス：年齢、アルツハイマー型認知症、および 2 型糖尿病。

社会への影響: 要介護度予測の実応用を想定した場合、地方自治体が公的な健康保険と介護保険の被保険者を網羅的かつ定期的に予測し、要介護の高リスク地域を把握するのに役立つと考えられる。日本の地方自治体は、運動、認知トレーニング、栄養指導などの介護予防プログラムを実施しており、高リスク地域を特定することは、これらのプログラムを実施する地域を選定するために有益である。