

【解説】

Preventive Medicine Reports

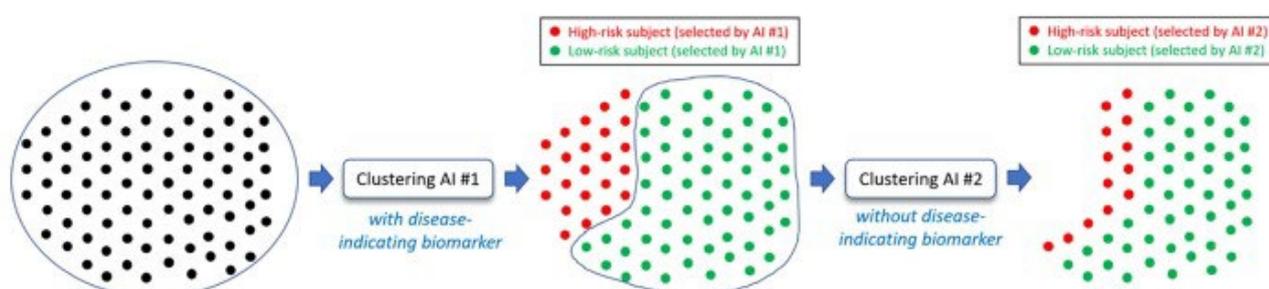
Volume 32, April 2023, 102129

Risk prediction of chronic diseases with a two-stage semi-supervised clustering method

二段階半教師ありクラスタリング法による慢性疾患のリスク予測

https://www.sciencedirect.com/science/article/pii/S2211335523000207?ref=pdf_download&fr=RR-9&rr=7ac0c9bb7977f6b5

この研究の目的は、疾患を示すバイオマーカーの測定値が正常範囲にある場合でも、将来の疾患のリスクを検出する方法を開発することです。



最初のクラスタリング AI によって選択された低リスク群の被験者について、疾患を示すバイオマーカーを含まない 2 番目のクラスタリング AI を使用して、高リスク被験者と低リスク被験者をさらに識別します。疾患を示すバイオマーカーを分析から除外することにより、このステップでは、疾患を示すバイオマーカーの値が正常範囲内にあるが、病気の患者と同様の他の測定値を持つ被験者をキャプチャできます。

この研究で使用されるデータは、疾病管理予防センター (CDC) の国民健康栄養調査 (NHANES) からのものです。2013 ~ 2014 年の NHANES (疾病管理予防センター、2013 年) の結果は、リスク評価アルゴリズムのトレーニングと開発に使用されました。別の 2011 ~ 2012 年の NHANES 調査の結果 (疾病管理予防センター、2012 年) が検証に使用されます。

心血管疾患 (CVD) や糖尿病などの慢性疾患の早期発見は、生死を分ける可能性があります。以前の研究

では、機械学習と疾患を示すバイオマーカーを使用した疾患の診断と予測の実現可能性が実証されています。この研究の目的は、疾患を示すバイオマーカーの測定値が正常範囲にある場合でも、将来の疾患のリスクを検出する方法を開発することです。

心血管（CVD）リスク評価

CVD の SKK-Means クラスタリングの結果を図 5 に示します。クラスタリングは、データを高リスクグループと低リスクグループに分類し、年齢と疾患の有病率に基づいて将来のリスクを予測します。心血管疾患の有病率は年齢とともに増加しますが、30 歳以上の人の高リスクグループと低リスクグループの間には有意な差があります。これは、20 代のハイリスクグループの人々が 40 代のときに心臓関連の問題を発症するリスクが 2 倍以上になる可能性があることを示しています。図 5 の数値は付録に記載されています。

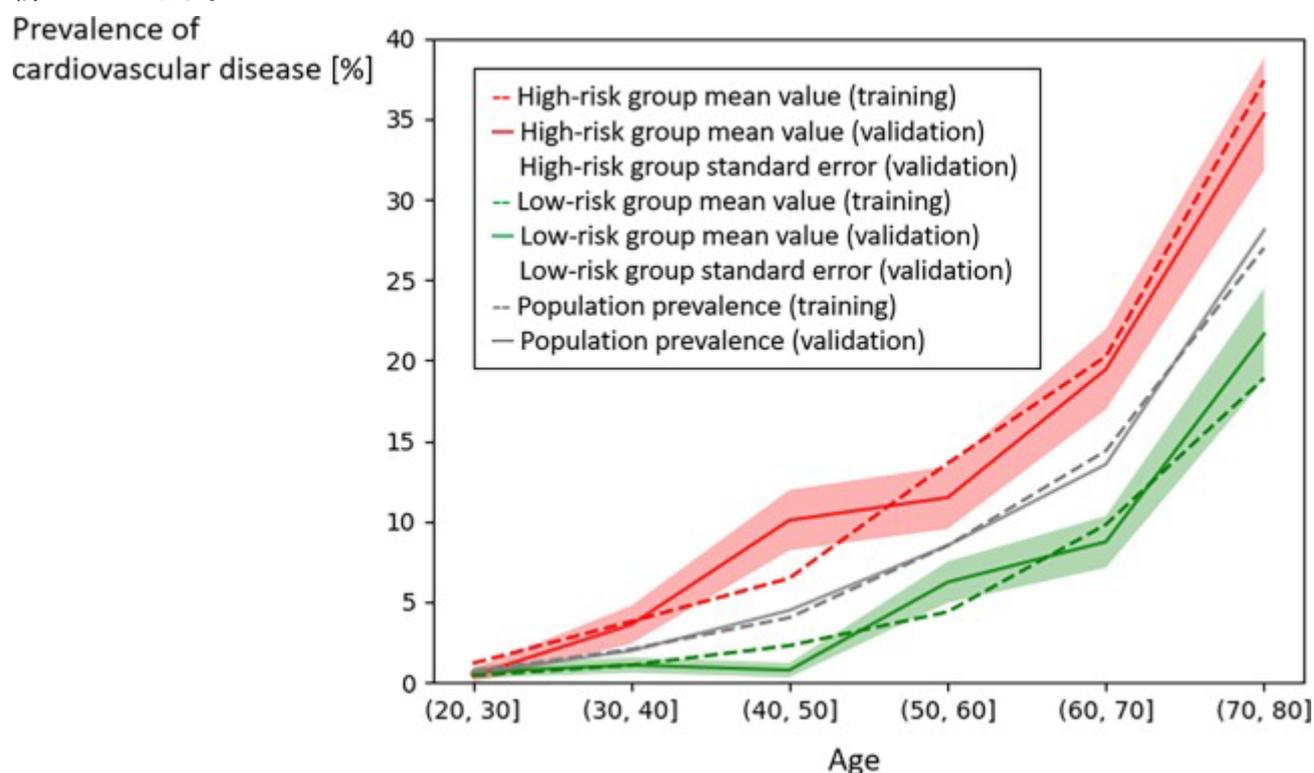


図 5. 年齢に対する CVD の有病率。赤い破線、トレーニング データの高リスクグループの平均値。緑色の線、トレーニング データの低リスクグループの平均値。赤線、検証データの高リスク群の平均値。赤色のバンド、検証データの高リスクグループの標準誤差。緑色の線、検証データの低リスクグループの平均値。緑の帯、検証データの低リスクグループの標準誤差。ラベル付けされたデータの 10% は、SKK-Means クラスタリングのグラウンド トゥールズとして使用されます。(この図の凡例における色への言及の解釈については、読者はこの記事の Web バージョンを参照してください。)

糖尿病リスク評価

糖尿病の有病率と年齢の結果を図 6 にまとめます。心血管疾患と同様に、糖尿病の有病率も年齢とともに上昇傾向を示しています。示されているように、高リスクグループの 30 歳以上の個人は、低リスクグループの個人と比較して、糖尿病を発症するリスクが少なくとも 2 倍あります。

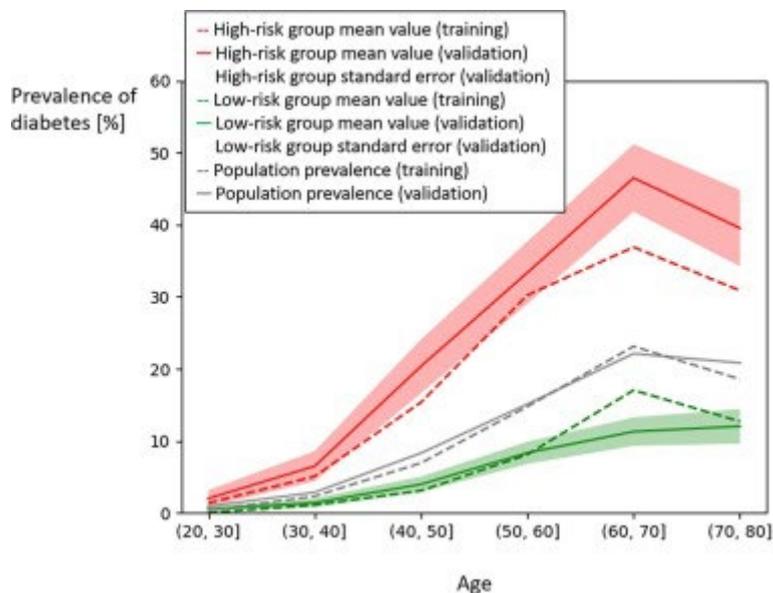


図 6. 糖尿病の有病率と年齢。図 5 と同じ配色。

議論

この研究では、心血管疾患と糖尿病の将来のリスクを検出するための 2 段階の半教師付きクラスタリング手法の実現可能性を示しました。最初の段階は、被験者の疾患を示すバイオマーカーがすでに異常なレベルにある場合に、疾患を示すバイオマーカーを使用して検出感度を最大化するために開発されました。第 2 段階は、疾患を示すバイオマーカーの影響を意図的に無視し、他のバイオマーカーに注目することで、偽陰性の結果をさらに減らすように設計されました。病気のリスク検出のパフォーマンスは、公開されている NHANES データセットを使用して評価および検証されます。

私たちの知る限りでは、これは将来のリスク評価のためにこのような多段階の半教師付きアプローチを取り入れた最初の研究です。

この研究では、リスクグループを分類するためのパラメーターは、個人がライフスタイルを維持した場合、大幅に変化しないと想定されていました。これを確実にするために、パラメーターは年齢依存性を取り除くために年齢性別調整されました。この仮定に基づいて、私たちの研究は、横断的データを使用して将来の疾患リスクを予測する可能性を実証しました。

将来的には、分類方法の有効性を評価し、関連するパラメーターを改善することで個人が病気のリスクを減らすことができるかどうかを判断するための縦断的研究を実施する予定です。成功すれば、これは全人類の健康管理に利益をもたらす可能性があります。