

子ども健康と環境に関する全国調査(エコチル調査)

論文概要の和文様式

雑誌における論文タイトル:

Decision tree learning for predicting chronic postpartum depression in the Japan Environment and Children's Study

和文タイトル:

慢性産後うつの決定木学習: エコチル調査

ユニットセンター(UC)等名: 富山ユニットセンター

サブユニットセンター(SUC)名:

発表雑誌名: Journal of Affective Disorders

年: 2024 DOI: 10.1016/j.jad.2024.10.034

筆頭著者名: 松村 健太

所属 UC 名: 富山ユニットセンター

目的:

これまで、多くの研究で機械学習の手法を用いた産後うつの予測モデルが構築されてきた。しかし、地域社会の母子保健の現場で紙とペンを使って簡単に実施できるほどシンプルなモデルには、ほとんど注目が集まっていなかった。そこで本研究では、決定木を用いた機械学習を実施し、慢性産後うつの予測モデルを構築した。

方法:

エコチル調査に参加する 84,091 人の母親を対象とした。産後 1 ヶ月及び産後 6 ヶ月の両時点でエジンバラ産後うつ尺度のスコアが 9 以上の場合に慢性産後うつと定義した。学習用データセットには、アンケートへの回答、健康状態、バイオマーカーなど、妊娠中に測定された 84 項目の多様な変数が含まれていた。学習においては、分岐の深さを 3 に、分岐ごとの枝の数を 4 に、枝ごとの n の最小数を 100 に制限した。学習用データと検証用データの比率は 7 対 3 とした。

結果:

35 の枝を持つ決定木が作成され、その受容動作特性曲線下面積は 0.84 であった。84 項目の変数のうち 10 項目が抽出され、分類に最も効果的だったのは「無価値感」であった。学習 ($n=58,635$) では、最も有病率の高い枝と最も有病率の低い枝はそれぞれ 73.2%と 0.84% (平均=6.29%) であり、検証 ($n=25,456$) ではそれぞれ 60.4%と 0.72% (平均=6.52%) であった。検証時の曲線下面積は、学習時と同等の 0.84 であった。

考察(研究の限界を含める):

本研究では、最大深度 3、分岐数 4、使用質問数 10 という、シンプルではあるものの、曲線下面積が 0.84 という比較的高性能な予測モデルを作成することに成功した。また、検証データセットの結果より、過学習(訓練データに適応しすぎて新しいデータでの予測精度が低下する現象)が生じていないことも示された。本研究の限界点としては、慢性産後うつの評価に自己記入式質問紙を使っていること、慢性産後うつの定義を産後 6 ヶ月までに基づいて行っていることである。

結論:

本研究では、シンプルで比較的高性能な予測モデルを作成した。このモデルは機械学習の専門知識がなくても容易に理解でき、使用できるため、草の根の地域保健を含む母子の健康の現場で役立つことが期待される。