PREDICTING VEGETABLE AND FRUIT CONSUMPTION IN GROUPS WITH DIFFERENT SOCIAL AND MATERIAL DEPRIVATION LEVELS USING MACHINE LEARNING

Mélina Côté, Catherine Laramée, Annie Lapointe, Simone Lemieux, Sophie Desroches, Ariane Bélanger-Gravel, Sharon Kirkpatrick, Benoît Lamarche

Understanding the applicability of predictive algorithms developed using machine learning (ML) to diverse groups, including those affected by structural disadvantage, is critical to pursuing nutrition equity. This study aimed to verify the applicability of a ML algorithm developed to predict adequate vegetable and fruit consumption (VFC) in a general population to subgroups with different social and material deprivation levels. Data from a sample of 2836 adults (86% women) from the NutriQuébec project included information on 94 variables potentially associated with dietary patterns. Material deprivation (MatDep; deprivation of goods and conveniences) and social deprivation (SocDep; deprivation of strong social networks), which reflect sociodemographic inequities, are categorized into quintiles (Q1= low deprivation, Q5= high deprivation) and were assigned for all participants using residential postal codes. The sample was randomly divided into training (60%), testing (20%) and validation (20%) sets. The outcome was adequate VFC (≥5 servings/d), measured by averaging intake across two to three web-based 24-h dietary recalls. The training set (general population) was used to develop a ML classification algorithm to predict adequate VFC. The model's prediction performance was evaluated in the testing set (general population) with the accuracy score (proportion of correct predictions). The algorithm was then applied to validation sets, divided into subgroups with low MatDep (Q1 and Q2) and high MatDep (Q4 and Q5) and low SocDep (Q1 and Q2) and high SocDep (Q4 and Q5), respectively. The classification algorithm developed in the training set predicted adequate VFC with an accuracy of 0.60 (95%CI:0.56-0.64) in the testing set. The model predicted adequate VFC in low and high MatDep subgroups with accuracies of 0.57 (95%CI:0.54-0.60) and 0.65 (95%CI:0.61-0.69), respectively. Similar results were obtained in the SocDep subgroups. The ML algorithm predicted adequate VFC with acceptable and similar accuracy in the high MatDep and SocDep subgroups as in the overall dataset, in which the model was developed and tested. These results suggest that ML models developed to predict a nutrition-related outcome in a general population could be applicable in subgroups of the population affected by material and social deprivation. [Canadian Institute for Health Research, Fonds de recherche du Québec-Santé]