Estimating childhood obesity prevalence and socio-economic deprivation using machine learning

– Gavin Long, Georgiana Nica-Avram, John Harvey, Evgeniya Lukinova, Roberto Mansilla, Simon Welham, Gregor Engelmann, Elizabeth Dolan, Kuzivakwashe Makokoro, Michelle Thomas, Edward Powell, James Goulding

Project overview

Childhood obesity has become a critical public health challenge in the UK, with rates soaring to 23.4% by age 11, particularly affecting children in the most deprived areas. The links between deprivation, poor nutrition, and adverse health outcomes are well established, yet measuring nutritional insecurity at scale remains difficult.

Traditional deprivation metrics like the Index of Multiple Deprivation (IMD) are updated only every 4-5 years and do not incorporate dietary data, making it challenging to track changes over time or understand the effect of new policies.

Project highlights and lessons learned

- **Developed a cost-effective machine learning approach** to measure the nutritional deficiencies in populations at a large scale
- **Introduced a new metric** for measuring calories per pound (£) spent called Calorie-oriented purchasing (COP)
- **Demonstrated how research methods can be used to assess the impact of food insecurity policies** on already struggling and deprived communities

Results and impact

To address the gap in data generation and policy impact, researchers at N/LAB, University of Nottingham, partnered with The Co-operative Group UK to develop a machine learning approach using massive anonymised transactional data—covering 4 million members and 2.5 billion transactions over 30 months. The team matched over 10,000 food products to their nutritional content and engineered novel variables related to obesogenic diets.

Most notably, the researchers introduced a new metric called **Calorie-oriented purchasing (COP)**—measuring the total calories obtained per pound spent—which captures the behavioural tendency to maximise calories for money.

Lead researcher John Harvey explains:

"COP emerged as one of the most robust predictors of both deprivation and childhood obesity. Areas with high COP showed strong associations with extreme deprivation and elevated childhood obesity rates, suggesting that economic pressures drive households toward cheap, calorie-dense foods."

Through comparative assessment of machine learning approaches, tree-based models (Random Forest, XGBoost) achieved the best performance, with 88% accuracy for predicting the most deprived neighbourhoods and 79% accuracy for childhood obesity prevalence. The models also revealed that higher sales of soft drinks, cigarettes, and grains, alongside lower purchases of fish, fruit and vegetables, and ready-made meals, were strongly associated with deprivation and higher levels of childhood obesity.

Critically, this work demonstrates that grocery transaction data can provide near real-time insights into neighbourhood-level nutritional insecurity and health risks. Unlike traditional surveys, which are expensive and infrequent, transactional data offers a rapidly updatable alternative for monitoring deprivation and informing targeted interventions.

The findings have important policy implications: they could enable more efficient allocation of resources to struggling communities, support evaluation of local policy interventions, and inform retailer-led initiatives such as discounting healthy foods in at-risk areas.

Scaling impact

The research ultimately recommends that retailers adopt new measures like COP for national nutrition surveillance, helping ensure that funding support reaches those most in need.