Artificial neural network to predict the effect of obesity on the risk of tuberculosis infection

Alaa Badawi,1,2* Christina J. Liu,3 Anas A. Rihem,4 Alind Gupta5

1Public Health Risk Sciences Division, Public Health Agency of Canada, Toronto, ON, Canada
2Department of Nutritional Sciences, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
3Department of Pharmacology and Toxicology, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
4Ontario Tech University, Oshawa, ON, Canada
5Cytel Inc., Toronto, ON Canada

Supplement Materials

Supplement Table 1: Hyperparameters tuned for the artificial neural network models.

Supplement Table 2: Specificity and sensitivity for the baseline methods used to evaluate the performance of the artificial neural network predicting latent tuberculosis from obesity and related factors.

Supplement Figure 1: Receiver operating characteristic curves for predicting latent tuberculosis infection from the random decision forests method, the support-vector machines method and the logistic regression method used for baseline evaluation.
Supplement Table 1: Hyperparameters tuned for the artificial neural network models.

| Hyperparameter | Network¹ |
|----------------|----------|
|                | I        | II       | III      |
| Momentum       | 0.3      | 0.5      | 0.5      |
| Learning Rate  | $10^{-1}$| $10^{-1}$| $10^{-1}$|
| Decay          | 0.7      | 0.3      | 0.3      |
| Alpha          | 0.3      | 0.4      | 0.4      |
| Batch Size     | 512      | 512      | 512      |
| Epochs         | 500      | 500      | 500      |

¹The Networks are all age- and sex-matched and include either BMI alone (Network I), BMI and HbA1C (Network II) or BMI, HbA1C and education (Network III).
Supplement Table 2: Specificity and sensitivity for the baseline methods used to evaluate the performance of the artificial neural network predicting latent tuberculosis from obesity and related factors.

| Method | Network\(^1\) |   |   |
|--------|---------------|---|---|
|        | I  | II | III |
| RDF    | Specificity  | 70| 98| 100|
|        | Sensitivity  | 94| 94| 80 |
| SVMs   | Specificity  | 70| 75| 95 |
|        | Sensitivity  | 65| 74| 71 |
| LG     | Specificity  | 80| 98| 83 |
|        | Sensitivity  | 84| 94| 96 |

\(^1\)The Networks are all age- and sex-matched and include either BMI alone (Network I), BMI and HbA1C (Network II) or BMI, HbA1C and education (Network III). Values shown presented as percent (%).

Abbreviations: RDF, random decision forests; SVMs; support-vector machines; LG, logistic regression and ANN, artificial neural network.
**Supplement Figure 1:** Receiver operating characteristic curves for predicting latent tuberculosis infection from the random decision forests method, the support-vector machines method and the logistic regression method used for baseline evaluation.

The three trained artificial neural networks were all age- and sex-matched and were trained for BMI (a), BMI and HBA1C (b) and BMI, HBA1C and level of education (c). AUC, area under the receiver operating characteristic curve. Figures depict the performance and the AUC for the three baseline evaluation methods used, i.e., the random decision forests (RDF) method, the support-vector machines (SVMs) method and the logistic regression (LG) method.
