

What Next: Predicting Symptom Occurrence in Travellers Using the ITIT App

Nadja Hedrich¹, Lakmal Meegahapola^{1,2,3}, Thibault Lovey¹, Patricia Schlagenhaut^{1,4}

¹ University of Zürich, Epidemiology, Biostatistics and Prevention Institute, Department of Global and Public Health, Zürich, Switzerland, ² ETH Zurich, Department of Health Sciences and Technology, Zürich, Switzerland, ³ Swiss Institute of Bioinformatics, Lausanne, Switzerland, ⁴ WHO Collaborating Centre for Travellers' Health, Department of Global and Public Health, MilMedBiol Competence Centre, Zürich, Switzerland

Background:

Whether a traveller becomes ill depends on many factors. Using the data from the ongoing ITIT (Illness Tracking in Travellers) study, we examined whether and machine learning models could predict symptom occurrence based on data input by participants.

Materials and Methods:

We analysed over 10,000 survey responses collected through the ITIT app, combining demographic, location, climate, and self-reported symptom data. Four machine learning models were trained to predict the occurrence of any symptom: (1) baseline risk using demographic and destination data, (2) next-day risk, (3) next-week risk, and (4) risk at any point during the trip. Nested cross-validation, hyperparameter tuning, and SMOTE oversampling were applied, and model performance was evaluated using AUROC and diagnostic plots. Model personalisation was done by weighting past data and incorporating participants' earlier responses to the prediction.

Results:

The baseline model alone showed limited performance (AUROC = 0.61). The next day model had a slightly improved AUROC at 0.65, increasing to 0.69 with personalisation. Next week prediction performed well, with at AUC of over 0.8 for both the RandomForest and XGBoost models. Finally, the models predicting symptom expression at any future point in the trip had an AUROC of 0.80, rising to 0.98 with personalised weighting. The top features impacting the models were not consistent over different methods, however Important predictors included travel duration, air quality, age, and continent of travel.

Conclusion:

Machine learning models demonstrated good predictive performance for travel-related symptom occurrence, with substantial improvement through personalised modelling.