Self-supervision of wearable sensors time-series data for influenza detection

evidation

Arinbjörn Kolbeinsson, Piyusha Gade, Raghu Kainkaryam, Filip Jankovic & Luca Foschini

Evidation Health, Santa Barbara, CA

Abstract

- Self-supervised learning may boost model performance in downstream tasks
- However, there is no principled way of selecting the self-supervised objectives that yield the most adaptable models
- We first show that self-supervised learning to predict next-day time-series values allows us to learn rich representations which can be adapted to perform accurate influenza-like illness (ILI) prediction
- Second, we perform an empirical analysis of three different self-supervised objectives to assess their adaptability to ILI prediction

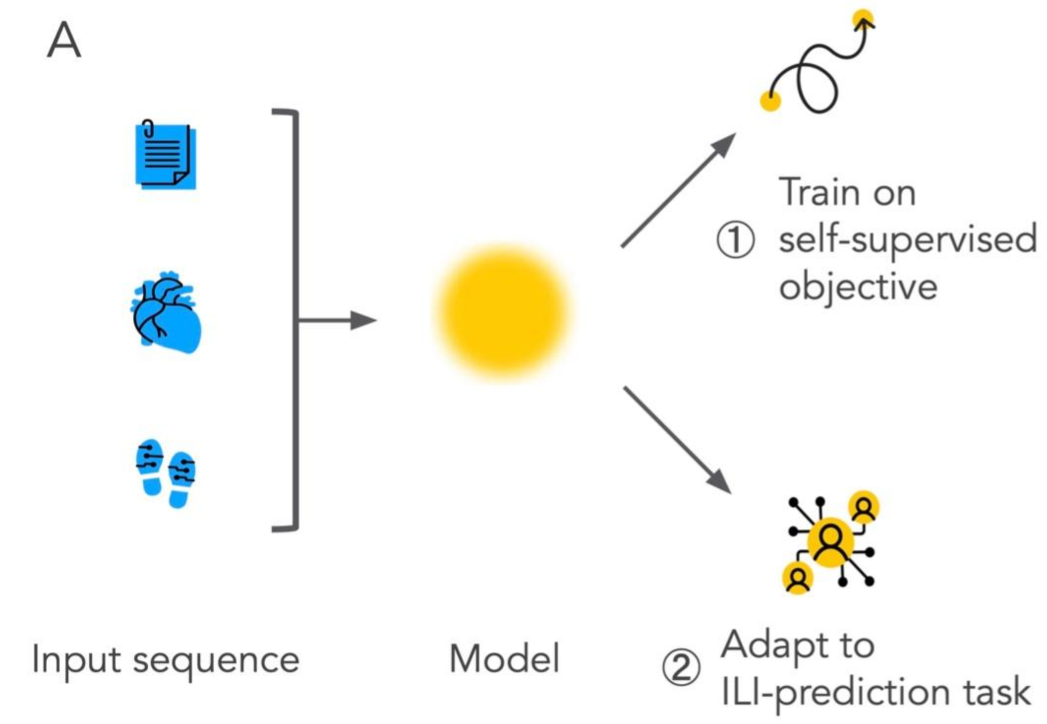
Data and set-up

We use the HomeKit dataset collected by Evidation Health between February and May 2020. 5229 individuals took part in a prospective cohort study where wearable (Fitbit) device data (including resting heart-rate, step count, and sleep) were recorded.

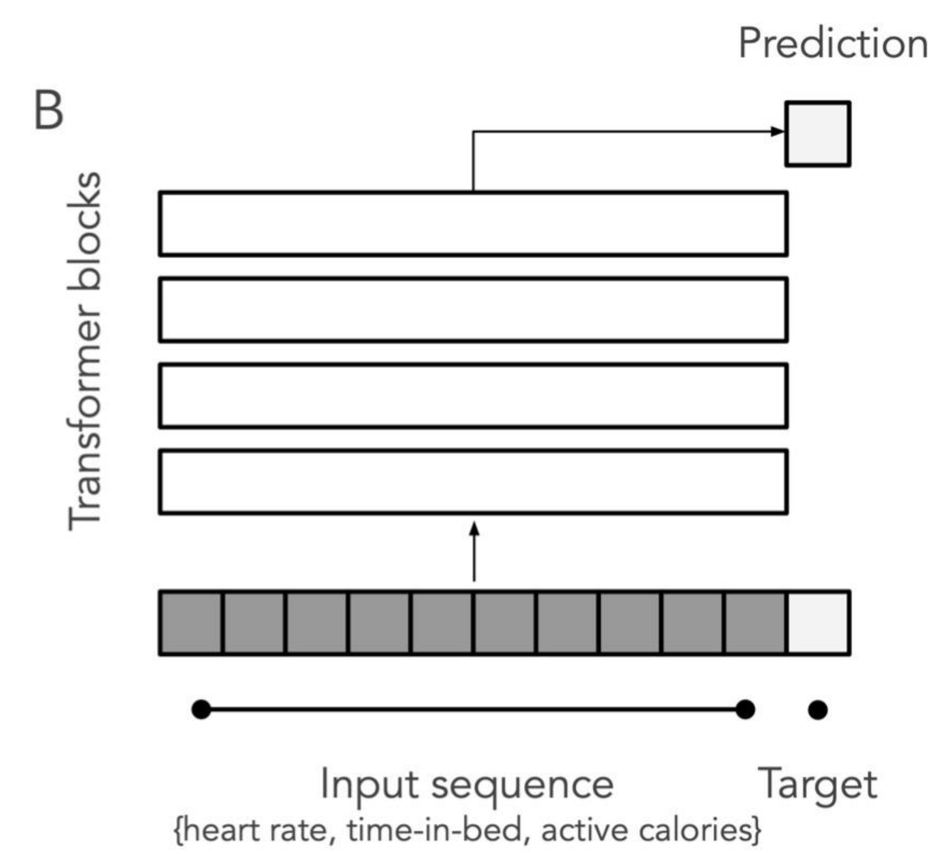
Every day, each participant was prompted to complete a survey on ILI-related symptoms. If a participant reported two symptoms on the same day, a nasal-swab was sent to a lab for PCR analysis to test for a viral infection.

For self-supervised learning, we train three models to predict next-day resting heart rate, time-in-bed minutes and activity calories expended. For evaluation, we apply the model to flu prediction. See figure A.

Model and learning



We use a decoder-only transformer-based architecture, inspired by state-of-the-art natural learning processing approaches. It is composed of four transformer blocks containing dense self-attention.

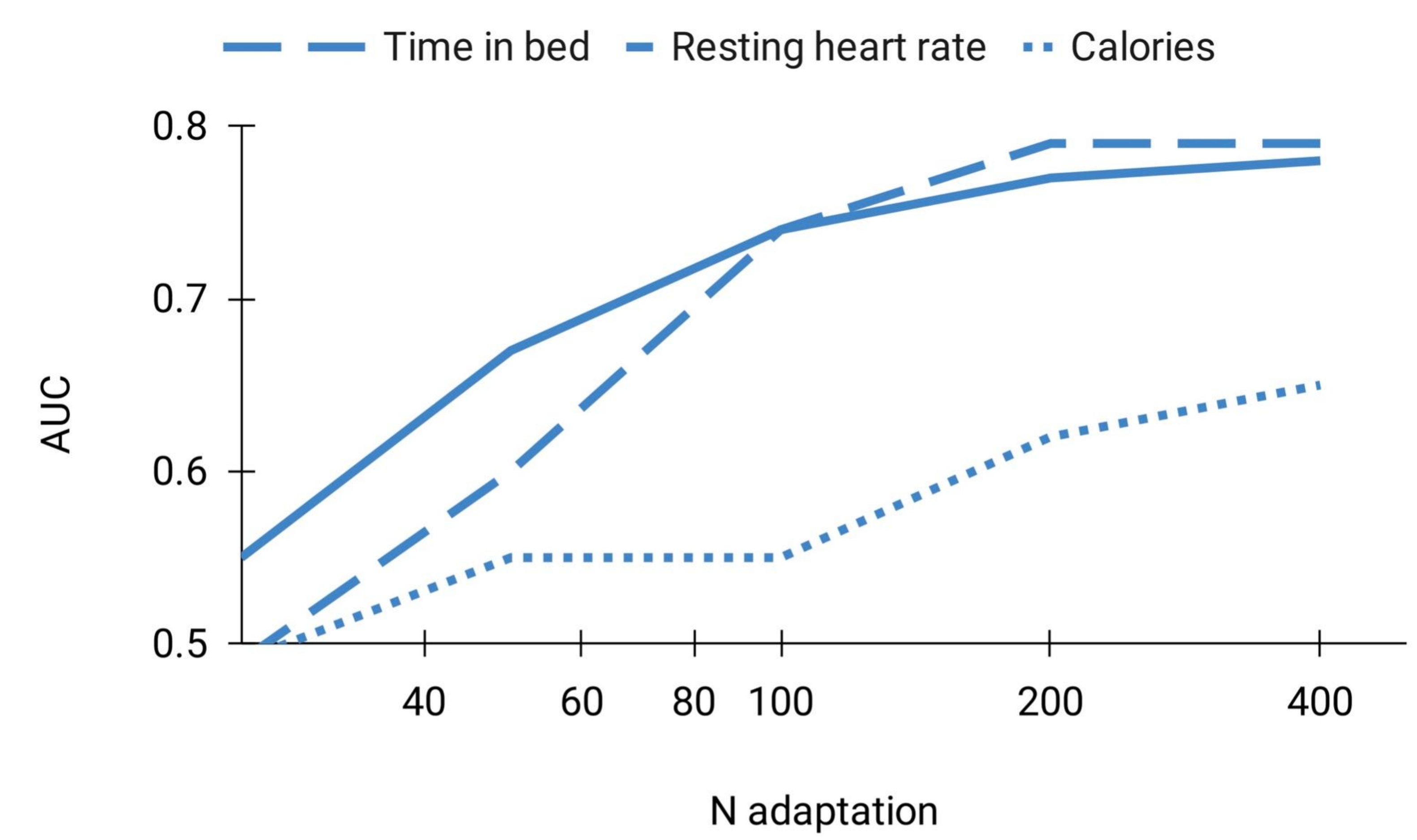


The final layer maps the output of the last transformer block to the task-dependent prediction. In regression tasks, this is a fully-connected layer with a single output unit. In the case of binary classification, the output is a vector of length two that, when combined with softmax activation, represent the predicted class pseudo-probabilities. See figure B.

Results

Out of the three models, the self-supervised objective on resting-heart rate and sleep gave the strongest result on the downstream task. With a small number of adaptation samples, the resting-heart rate model achieves better performance than the other two models.

With more adaptation samples, the model adapted from time-in-bed eventually matches the one trained on resting-heart rate.



Discussion

We showed that a model self-trained on predicting next-day resting heart-rate or sleep outperforms a model trained using calories expended on a downstream ILI task. Task relatedness remains an active research problem. Future work includes learning all features together and, after further tuning, compare against traditional baselines and perform more validation.