Detection of COVID-19 using multimodal data from a wearable device: results from the frst TemPredict Study

In this paper researchers proposed wearable devices that continuously measure physiological metrics hold promise as tools for early illness detection. Authors developed algorithm that identifed COVID-19 an average of 2.75 days before participants sought diagnostic testing with a sensitivity of 82% and specifcity of 63%. ROC and AUC was 0.819 including continuous temperature yielded an AUC 4.9% higher than without this feature.

Consumer wearable devices that continuously measure physiological metrics such as dermal temperature, heart rate, and respiratory rate can establish users individual baseline patterns and allow detection of deviations from their baselines.

Data availability: Oura's data use policy does not permit us to make the data available to third parties without approval. This application process will require requesters to make a written commitment expressing agreements to not duplicate data, to not share data with third parties, and/or other confidentiality precautions.

Researchers used machine learning methods to train a classifier algorithm using physiological data to distinguish participants during a period around their diagnosis of COVID-19 versus during a comparison period from the same participants about six weeks prior to developing COVID-19. They successfully trained classifier models on three-day windows relative to PX. They identified three key dates that would allow algorithmic training and performance: traditional diagnosis date (DX), initial symptom reporting date (SX) and a novel date of physiological alterations(PX). Authors developed an algorithm to identify COVID-19 onset using data collected by a commercially available wearable device. The resultant algorithm had high sensitivity (82%), with moderate specifcity (63%). In developing this algorithm, they placed greater emphasis on sensitivity than on specificity, as their goal was to develop an algorithm that could efectively identify individuals who should obtain laboratory-based diagnostic testing.

Methods: Researchers recruited adults from the broader population who already possessed Oura Rings by sending them invitations within the Oura App on their smartphones. Prospective participants could tap on this invitation, which linked to the UCSF consent survey online.They did not have a recruitment ceiling for participants who met eligibility criteria and possessed their own Oura Rings. They recruited frontline healthcare workers at participating sites by enlisting leadership at each institution and obtaining IRB review at each institution.

Measures: Baseline self-report survey, Daily self-report surveys,Monthly self-report surveys, Dried blood spot (DBS) antibody testing, Oura ring data were collected.

Variable creation: Authors created several variables for analysis. It consists Diagnosis determination (DX), Confirmed positive cases who reported a positive COVID-19 test result on an oral or nasopharyngeal swab, saliva, stool, or antigen test, Confrmed negative cases s who tested negative on study-provided DBS antibody testing and who did not report positive COVID-19 test results in any study survey, Test ambiguous cases r who self-reported a negative antibody test result afer a reported positive swab, saliva, antigen, or stool test ,Survey ambiguous cases,DX-generated case lists, DX,SX and PX region.

Data preparation: Te Oura Ring records five physiological metrics (data streams) on the scale of minutes. Authors aggregated data from each of the five streams within 30-min, consecutive time

intervals that overlapped by 15 min. Authors chose these time frames to balance computational resolution.

Algorithmic description: Authors created a machine learning pipeline that detected physiological features distinguishing COVID-19 illness from non-illness. This pipeline had 3 constituent parts:
1) data processing module,
2) short-time classification and detection module,
3) post-detection "trigger" logic module.

The data processing steps were to a) gather and b) normalize individuals' data. Authors trained a set of five Random Forest models on the normalized data sketches and trend time series. The classifier training samples consist of data from overlapped 30-min intervals from individuals assigned to the training set and each of the five models were differentiated by considering distinct time frames as the positive class. The set of trained classifier models were then used to predict a preliminary score at each interval assessing. The five Random Forest models were trained such that each model encompassed a different positive time frame near PX. The negative training samples were held constant for all five models. The first of these models was trained on data sketch and trend variable values drawn from the range PX − 3 to PX − 1, the second covered PX − 2 to PX,the third, PX − 1 to PX+1, fourth, PX to PX+2, and fifth PX+1 to PX+3. In this way, authors learned patterns relevant to infection at each of several "early-stage" time frames in the vicinity of illness onset.

Performance evaluation: The detection performance of our pipeline was evaluated via a five-fold cross validation using data from the identified training cohort (n=73). Authors calculated ROC curves and their corresponding AUC using the short-time detection scores.