Evaluation of Machine Learning to Detect Influenza Using Wearable Sensor Data and Patient-Reported Symptoms: A Cohort Study

Kamran Farooq, Melody Lim, Lawrence Dennison-Hall, Finn Janson, Aspen Hazel Olszewska, Muhammad Mamduh Ahmad Zabidi, Anna Haratym-Rojek, Karol Narowski, Barry Clinch, Marco Prunotto, Devika Chawla, Victoria Hunter, Vincent Ukachukwu

Submitted to: Journal of Medical Internet Research
on: April 05, 2023

Table of Contents

Evaluation of Machine Learning to Detect Influenza Using Wearable Sensor Data and Patient-Reported Symptoms: A Cohort Study

Kamran Farooq[1]; Melody Lim[2]; Lawrence Dennison-Hall[3]; Finn Janson[3]; Aspen Hazel Olszewska[4]; Muhammad Mamduh Ahmad Zabidi[5]; Anna Haratym-Rojek[4]; Karol Narowski[6]; Barry Clinch[3]; Marco Prunotto[7, 8]; Devika Chawla[2]; Victoria Hunter[2]; Vincent Ukachukwu[3]

[1]Roche Data & Analytics Chapter (Data Science) Kaiseraugst CH
[2]Genentech, Inc. South San Francisco US
[3]Roche Products Ltd Welwyn Garden City GB
[4]Roche Global IT Solution Centre Warsaw PL
[5]Roche Services (Asia Pacific) Sdn. Bhd. Subang Jaya MY
[6]Badger Software Sp. z o.o. Wroclaw PL
[7]Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva Geneva CH
[8]F. Hoffmann-La Roche Ltd Basel CH

Corresponding Author:
Kamran Farooq
Roche Data & Analytics Chapter (Data Science)
Wurmisweg
Kaiseraugst
CH

Abstract

Background: Machine learning offers quantitative pattern recognition analysis of wearable device data and has the potential to detect illness onset and monitor influenza-like illness (ILI) in infected patients.

Objective: To evaluate the ability of machine learning algorithms to distinguish between influenza-positive and influenza-negative participants in a cohort of symptomatic ILI patients using wearable sensor (activity) data and self-reported symptom data during the latent and early symptomatic period of ILI.

Methods: This cohort study used the extreme gradient boosting (XGBoost) classifier to determine whether a participant was influenza positive or negative based on three models: symptoms data only; activity data only; and combined symptoms and activity data. Data were collected from the Home Testing of Respiratory Illness (HTRI) study and FluStudy2020, both conducted between December 2019 and October 2020. Analyses included participants in these studies with an at-home influenza diagnostic test result. Fitbit devices were used to measure participants' steps, heart rate, and sleep data. Participants detailed their ILI symptoms, healthcare-seeking behaviors, and quality of life. Model performance was assessed by: area under the curve (AUC), balanced accuracy, recall (sensitivity), precision (positive predictive value [PPV]), and F2 (weighted harmonic mean of precision and recall) score.

Results: An influenza diagnostic test result was available for 953 and 925 participants in HTRI and FluStudy2020 respectively, of whom 848 and 840 had activity data. The highest-performing model used symptoms and activity data (test set means: 0.74 AUC, 0.68 balanced accuracy, 0.70 sensitivity, 0.38 PPV, 0.60 F2). The symptoms-only model had the second-best performance (0.74 AUC, 0.69 balanced accuracy, 0.65 sensitivity, 0.42 PPV, 0.58 F2). The top features guiding influenza detection were cough, mean resting heart rate during main sleep, and fever for the combined model, and cough, fever, and chills for the symptoms-only model.

Conclusions: Machine learning algorithms had insufficient accuracy to detect influenza, suggesting that previous findings from research-grade sensors tested in highly controlled experimental settings may not easily translate with scalable commercial-grade sensors. In the future, more advanced wearable sensors may improve their performance in the early detection and discrimination of viral respiratory infections. Clinical Trial: NCT04245800

(JMIR Preprints 05/04/2023:47879)
DOI: https://doi.org/10.2196/preprints.47879

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original paper

Evaluation of Machine Learning to Detect Influenza Using Wearable Sensor Data and Patient-Reported Symptoms: A Cohort Study

Manuscript word count: 4256/6000

Abstract

Abstract word count: 421/450

Background: Machine learning offers quantitative pattern recognition analysis of wearable device data and has the potential to detect illness onset and monitor influenza-like illness (ILI) in infected patients.
Objectives: To evaluate the ability of machine learning algorithms to distinguish between influenza-positive and influenza-negative participants in a cohort of symptomatic ILI patients using wearable sensor (activity) data and self-reported symptoms data during the latent and early symptomatic period of ILI.
Methods: This prospective observational cohort study used the extreme gradient boosting (XGBoost) classifier to determine whether a participant was influenza positive or negative based on three models: symptoms-only data; activity-only data; and combined symptoms and activity data. Data were collected from the Home Testing of Respiratory Illness (HTRI) study and FluStudy2020, both conducted between December 2019 and October 2020. The model was developed using the FluStudy2020 data and tested on the HTRI data. Analyses included participants in these studies with an at-home influenza diagnostic test result. Fitbit (Fitbit LLC) devices were used to measure participants' steps, heart rate, and sleep data. Participants detailed their ILI symptoms, healthcare-seeking behaviors, and quality of life. Model performance was assessed by: area under the curve (AUC), balanced accuracy, recall (sensitivity), specificity, precision (positive predictive value [PPV]), negative predictive value (NPV), and weighted harmonic mean of precision and recall (F2) score.
Results: An influenza diagnostic test result was available for 953 and 925 participants in HTRI and FluStudy2020, respectively, of whom 848 and 840 had activity data.
For the training and validation sets, the highest performing model was trained on the combined symptoms and activity data (training AUC, 0.77; validation AUC, 0.74) vs symptoms-only (training AUC, 0.73; validation AUC, 0.72) and activity-only (training AUC, 0.68; validation AUC, 0.65). For the FluStudy2020 test set, model performance with the combined symptoms and activity data was closely aligned with the symptoms-only model (combined symptom and activity test AUC, 0.74; symptoms-only test AUC, 0.74). These results were validated using independent HTRI data (combined symptom and activity evaluation AUC, 0.75; symptoms-only evaluation AUC, 0.74). The top features guiding influenza detection were cough, mean resting heart rate (RHR) during main sleep, fever, and total minutes in bed for the combined model, and fever, cough, and sore throat for the symptoms-only model.
Conclusions: Machine learning algorithms had moderate accuracy to detect influenza, suggesting that previous findings from research-grade sensors tested in highly controlled experimental settings may not easily translate with scalable commercial-grade sensors. In the future, more advanced wearable sensors may improve their performance in the early detection and discrimination of viral respiratory infections.

Trial Registration: NCT04245800

Keywords: influenza; influenza-like illness; wearable sensor; person-generated health care data; machine learning

Introduction

Between 2010 and 2020, an estimated 9 to 41 million annual illnesses were attributed to influenza infection in the United States (US) [1]. Estimates of US annual hospitalizations ranged from 140,000 to 710,000 and deaths from 12,000 to 52,000 [1]. Early diagnosis and implementation of non-pharmaceutical interventions (eg, quarantine) are critical in preventing onward transmission and reducing disease burden of influenza-like illnesses (ILIs). Recent data show that wearable devices (eg, fitness trackers and smartwatches) may help to detect viral infection before symptoms develop and may provide an early warning system for viral illness [2-12].

In 2019 and 2020, approximately 29% of US adults were reported to use wearable devices [13], which range from fitness trackers that passively record heart rate (HR) and number of daily steps, to more sophisticated devices that can measure parameters such as sleep duration and quality, blood pressure, blood glucose, and oxygen saturation levels [14]. The wide availability and increasing popularity of wearable devices makes them convenient, passive tools to record person-generated health data that could be harnessed to improve both individual and public health.

Wearable devices have the potential to detect the onset of illness and monitor disease progression or severity in virus-infected patients. In the future, this may allow people to be alerted to possible infection or the need to seek medical care in the early stages of disease [2-6]. Grzesiak et al. have further shown that wearable devices may be able to predict the infection severity profile of a patient up to 24 hours before the onset of symptoms following exposure to influenza virus or rhinovirus [4]. The ability to predict illness severity may provide opportunities to discriminate between respiratory viral infections with more severe clinical presentation and which carry greater risk to public health, such as influenza and coronavirus disease 2019 (COVID-19), and those with a mostly mild clinical presentation, such as the common cold. Our previous analysis objectively characterized the *wearable phenotype* of individuals with ILI as well as those with confirmed influenza infection [15]. We demonstrated that before symptom onset, and throughout the duration of an ILI event, individuals experience reduced total daily steps, total active time, and sleep efficiency, as well as increased sleep duration and changes in resting heart rate (RHR) [15].

Here, we report the development and evaluation of a machine-learning model to detect laboratory-confirmed influenza infection based on wearable sensor and symptoms data in the latent and early symptomatic period (up to 1 day after symptom onset), using an extreme gradient boosting (XGBoost) classifier [16, 17] in a cohort of symptomatic ILI patients.

Machine learning offers a quantitative analysis of the data collected from wearable devices; XGBoost is an optimized distributed gradient-boosting library that implements machine-learning algorithms, providing parallel-tree boosting [17]. XGBoost, a supervised machine-learning process, can be used to solve classification tasks, in which one can determine if an instance is in a particular category by studying the features of that instance [17].

Using commercial wearable sensors (Fitbit), it has previously been demonstrated that nationwide mobility (measured as total daily steps in a US population) decreased due to ILI symptoms, and that ILI burden (determined by the difference in total daily steps) was associated with care-seeking behaviors, number of workdays missed, and self-reported overall health [18]. Another study showed

that abnormalities in RHR and sleep duration, measured by wearable sensors, could be leveraged to predict the real-time incidence of ILI [19]. Recently, wearable sensor data have also been used to assess physiological signs associated with COVID-19 [5, 10, 12, 20-26].

The objective of our analysis was to determine the ability of an XGBoost model to distinguish between influenza-positive and influenza-negative participants during the latent and early symptomatic period of ILI (Days –4 to +1). Wearable and symptoms data were used, gathered from two independent studies, FluStudy2020 and the Home Testing of Respiratory Illness (HTRI) study (NCT04245800); the former was used for training, testing, and validation, and the latter was used as a secondary holdout set for evaluation.

Methods

Study Design and Participants / Overview

This prospective observational cohort study evaluated the ability of machine learning algorithms to distinguish between influenza-positive and influenza-negative participants in a cohort of patients with ILI. Analyses were conducted using wearable sensor (activity) data and self-reported symptom severity data from participants enrolled in FluStudy2020, with an influenza diagnostic test result from a self-administered kit [15, 27]. Data from the HTRI study were used as an independent holdout set. All participants provided written consent.

The XGBoost model was used to classify whether a participant was influenza positive or negative based on three different models: symptoms-only data, activity-only data, and combined symptoms and activity data. Other models were not assessed with these data on the basis of previous internal analyses with different data, in which H2O AutoML was used to train and tune various models; XGBoost was found to be the best performing model. Participant variables, including age, gender, body mass index (BMI), and month in which the participant conducted an at-home influenza test, were considered before the Boruta feature selection algorithm was applied to the activity-only and combined symptoms and activity models. The symptoms-only model included all participant variables. The XGBoost model was assessed for its early detection of influenza infection in the FluStudy2020 training, validation, and test sets, as well as the HTRI secondary holdout set, using the following metrics: balanced accuracy, recall (sensitivity), specificity, precision (positive predictive value [PPV]), negative predictive value (NPV), weighted harmonic mean of precision and recall (F2) score, and area under the receiver operating characteristics curve (AUC ROC). Calibration plots and feature importance plots were generated for each of the three models. A model evaluation schematic is shown in Figure S1 in Multimedia Appendix 1.

Data Collection and Pre-processing

The HTRI study and FluStudy2020 were conducted by Evidation Health in adults in the United States (US) between December 2019 and October 2020. Participants in each study were aged ≥18 years, lived in the US, and owned and were willing to wear a Fitbit device during the day as well as during sleep for the duration of the study. Full inclusion and exclusion criteria are shown in Table S1 in Multimedia Appendix 1. Steps, HR, and sleep data were collected through continuous passive monitoring via the participants' Fitbit devices. Participants also completed daily surveys of whether they experienced flu symptoms in the past 24 hours, self-reported ILI symptom severity, healthcare-seeking behaviors, and quality of life. Biweekly and monthly surveys were used to capture influenza-related complication events and vaccination history. Participants reporting certain ILI symptoms were instructed to perform a self-administered influenza diagnostic test. Samples were returned to the laboratory for confirmation of influenza by a highly sensitive reverse transcription polymerase

chain reaction (RT-PCR) test. The primary assessment of data from the two studies, including the removal of physiologically implausible data or null estimates, has been described previously [15]. Missing data were automatically handled by XGBoost by finding the best split direction when missing data were present.

Participants with an influenza diagnostic test result were identified and activity data were assessed for quality and completeness for each participant-day. Steps data were considered valid if the participant had at least 10 hours of steps wear-time [28, 29], or if they had a valid HR day. HR data were considered valid if they included a minimum of 600 minutes (10 hours) of HR data and if a Fitbit-estimated RHR measure was available for that day. Sleep data were considered valid if non-zero and non-missing total sleep minutes were available for the day. Finally, any day with less than 10 hours of wear-time was considered invalid.

The maximum self-reported severity of eight symptoms was analyzed: early fever, sore throat, cough, headache, muscle ache, chills, fatigue, and nasal congestion. In total, 41 activity features were analyzed, including: RHR, total minutes asleep, total number of steps, proportion of the day that the participant spent being physically active (defined as ≥50 steps per minute), maximum amount of activity the participant was able to complete within a single hour of the day, sleep efficiency score during main sleep, minutes in bed for the main sleep only of the day, number of naps, total minutes in bed, percentage of minutes with HR >1.5× RHR for the day, proportion of minutes with non-zero steps out of the total minutes the device was worn, and mean RHR during main sleep. In addition, 29 HR variability (HRV) features were analyzed, derived from RHR captured during the participant's sleep period.

Model Building and Optimization

Baseline predictors were assessed based on including symptoms-only features, activity-only features, and then combining both features for the latent period to Day 1 of ILI (Days –4 to +1). This baseline model was built on the XGBoost classifier, which was selected as the machine-learning algorithm trained to detect influenza due to its scalability, regularization, and ability to detect complex non-linear relationships. Metric calculations were based on a previously published study [5]:

$$\frac{\text{mean (activity feature [test data]) – median (activity feature [baseline data])}}{\text{interquartile range (IQR) for all activity feature values}}$$

where [test data] represents ILI Days –4 to +1, which encompass the latent period (Days –4 to –1; ie, the incubation period for influenza), ILI onset (Day 0), and part of the early symptomatic period (Day +1), and [baseline data] represents the participants' healthy baseline data from 2 weeks prior to the latent period (Days –18 to –5).

The model was optimized using the Bayesian hyperparameter optimization algorithm, a Bayesian inference and Gaussian process to find the maximum value of an unknown function with minimal iterations. AUC ROC was the metric subject to optimization; 100 optimization trials were run per model (symptoms-only, activity-only, and combined symptoms and activity data; each model included participant variables). The parameters optimized were: maximum number of trees: (2, 50), learning rate: (0.0001, 0.2), maximum tree depth: (2, 10), subsample ratio of training instances prior to growing trees: (0.2, 1.0), column subsample ratio at each level: (0.1, 1.0), column subsample ratio at each tree: (0.1, 1.0), column subsample at each node: (0.1, 1.0), maximum delta step allowed by each leaf output: (0, 10), minimum sum of instance weight needed in a child (sub-tree: [0.0, 10.0]), L1 regularization: (0.00001, 1) and L2 regularization: (0.00001, 100). To mitigate the imbalanced ratio of influenza-positive to

influenza-negative participants, class weights were calculated and applied to the model to give greater weight to the minority influenza-positive class.

Given these differences in study design, the current analysis excluded participants who tested positive for influenza in the HTRI study but did not meet any of the FluStudy2020 ILI population criteria. There was also a small subset of HTRI participants whose symptoms would have met the influenza test kit criteria for FluStudy2020, but from whom a sample was not collected because they did not meet the HTRI influenza test kit criteria. To combine the HTRI dataset with FluStudy2020 data, it was necessary to verify that the HTRI participants' self-reported illness dates (which were provided in the same recovery survey in which a healthcare visit was reported) aligned with the analysis-derived ILI event dates (which were created during analysis of the daily survey responses). This permitted verification that only healthcare visits during the same illness period as the ILI event period were included in the analysis. HTRI participants were only categorized as having made or not made a healthcare visit if their self-reported ILI event period overlapped with the analysis-derived ILI event period.

Model Validation and Evaluation

Stratified k-fold shuffle cross validation (k=50) was used to ensure model performance was reliable and robust. Overall, 50 models were trained on different training/validation sets before being evaluated on a single test set and the HTRI data. Of the FluStudy2020 data, 64% were used for k-stratified splits consisting of 50 training sets, and 16% were used for validation. The remaining 20% of the data were set aside as the testing set. External validation was performed using the HTRI data. The Boruta feature selection algorithm was applied to reduce the dimensions of the activity features to minimize the impact of noise and reduce overfitting.

The model performance was assessed by the following metrics: balanced accuracy, recall (sensitivity), specificity, precision (PPV), NPV, F2 score, and AUC ROC. The model performance results consisted of the AUC ROC curves and mean performance across each k-fold along with 95% confidence intervals (CIs) for the training, validation, and test sets. The distribution of positive and negative predictions for the aggregated performance (based on symptoms features, activity features, and these features combined) were described using confusion matrices. The values in each confusion matrix comprised the mean across each fold with their respective 95% CIs. Feature importance analyses were performed for each model, with the most important features summarized in feature importance plots.

Software

Analyses were performed using Python version 3.7; *xgboost 1.5.2* was used for modeling and *bayesian-optimization 1.2.0* was used for hyperparameter optimization. Feature importance was determined using XGBoost built-in feature importance. Data processing and visualization was performed with *pandas 1.3.4*, *NumPy* 1.21.4, and *Matplotlib* 3.5.1 [30-32]. *Kedro* was used to build robust and scalable data pipelines [33]. Feature selection was performed with *Boruta 0.3* [34]. Statistical analysis was performed with *SciPy 1.7.3* [35]. Metric computation, k-fold data splitting, and class weight calculations were performed using *scikit-learn 0.24.2* [36]. The Python package *hrvanalysis* [37] was used to derive HRV features.

Ethical Considerations

The HTRI study and FluStudy2020 were conducted by Evidation Health, Inc. Institutional review

board (IRB) approval was given by WCG for both the HTRI study (study number: 1271380; tracking number: 20192965) and FluStudy 2020 (study number: 1271500; tracking number: 20193003). Participants were recruited from the Evidation consumer platform, a free application that allows members to earn compensation for completing surveys, sharing health activity data, and reading health articles. Individuals were given the opportunity to enroll into the study by providing informed consent to complete study activities and for use of their data. Participants earned reward points, redeemable for money, as compensation for completing study activities. Reward points worth up to $10 were available on completion of enrollment and a maximum of $109 dollars could be earned over the course of the study, if all study activities were completed. The data used for analysis were de-identified; each subject enrolled in the study was coded with a unique subject identification number.

Results

Participants

FluStudy2020 had 925 participants, of whom 840 had activity data that met the data density criteria. Of these, 639 were influenza-negative and 201 were influenza-positive (Figure S1 in Multimedia Appendix 1). The HTRI study had 953 participants, and activity data meeting the data density criteria were available for 848 participants. Of these, 657 were influenza-negative and 191 were influenza-positive. Baseline demographics of participants included in the model evaluation are presented in Table 1. Most participants were female (91.0% and 77.8%) with mean (standard deviation [SD]) ages of 37.4 (9.6) and 37.6 (9.1) for FluStudy2020 and HTRI, respectively. Distributions of age, BMI, and gender were balanced between the influenza-negative and influenza-positive groups. The maximum self-reported symptom severities and wearable sensor data during ILI Days –4 to +1 are shown in Table 2 and Table 3, respectively.

Table 1. Baseline demographics of participants included in the model evaluation.

	FluStudy2020			**HTRI**[a]		
Characteristics	Overall (n=840)	Influenza negative (n=639)	Influenza positive (n=201)	Overall (n=848)	Influenza negative (n=657)	Influenza positive (n=191)
Age, mean (SD[b])	37.42 (9.59)	37.10 (9.35)	38.45 (10.25)	37.55 (9.10)	37.47 (9.15)	37.83 (8.95)
BMI[c], mean (SD)	31.25 (8.16)	31.42 (8.16)	30.72 (8.18)	30.64 (7.51)	30.64 (7.69)	30.61 (6.89)
Region, n (%)						
Midwest	291 (34.64)	206 (32.24)	85 (42.29)	299 (35.26)	216 (32.88)	83 (43.46)
Northeast	139 (16.55)	109 (17.06)	30 (14.93)	134 (15.80)	99 (15.07)	35 (18.32)
South	257 (30.60)	198 (30.99)	59 (29.35)	239 (28.18)	191 (29.07)	48 (25.13)
West	153 (18.21)	126 (19.72)	27 (13.43)	176 (20.75)	151 (22.98)	25 (13.09)
Sex, n (%)						
Female	764 (90.95)	586 (91.71)	178 (88.56)	660 (77.83)	519 (79.00)	141 (73.82)
Male	71 (8.45)	48 (7.51)	23 (11.44)	186 (21.93)	136 (20.70)	50 (26.18)
Non-binary	5 (0.60)	5 (0.78)	0 (0.00)	2 (0.24)	2 (0.30)	0 (0.00)
Race, n (%)						
Asian	16 (1.90)	13 (2.03)	3 (1.49)	27 (3.18)	19 (2.89)	8 (4.19)
Black or African American	31 (3.69)	25 (3.91)	6 (2.99)	21 (2.48)	14 (2.13)	7 (3.66)
Multiple races	32 (3.81)	27 (4.23)	5 (2.49)	28 (3.30)	23 (3.50)	5 (2.62)
Alaska Native, American Indian, Native Hawaiian or other Pacific Islander	2 (0.24)	2 (0.31)	0 (0.00)	1 (0.12)	1 (0.15)	0 (0.0)
Other	6 (0.71)	4 (0.63)	2 (1.00)	8 (0.94)	7 (1.07)	1 (0.52)
White	753 (89.64)	568 (88.89)	185 (92.04)	763 (89.98)	593 (90.26)	170 (89.01)

[a]HTRI: Home Testing of Respiratory Illness.
[b]SD: standard deviation.
[c]BMI: body mass index.

Table 2. Frequency of maximum self-reported symptom severity during ILI[a] Days –4 to +1.

	FluStudy2020			**HTRI[b]**		
Characteristics	Overall (n=840)	Influenza negative (n=639)	Influenza positive (n=201)	Overall (n=848)	Influenza negative (n=657)	Influenza positive (n=191)
Age, mean (SD[c])	37.42 (9.59)	37.10 (9.35)	38.45 (10.25)	37.55 (9.10)	37.47 (9.15)	37.83 (8.95)
BMI[d], mean (SD)	31.25 (8.16)	31.42 (8.16)	30.72 (8.18)	30.64 (7.51)	30.64 (7.69)	30.61 (6.89)
Region, n (%)						
Midwest	291 (34.64)	206 (32.24)	85 (42.29)	299 (35.26)	216 (32.88)	83 (43.46)
Northeast	139 (16.55)	109 (17.06)	30 (14.93)	134 (15.80)	99 (15.07)	35 (18.32)
South	257 (30.60)	198 (30.99)	59 (29.35)	239 (28.18)	191 (29.07)	48 (25.13)
West	153 (18.21)	126 (19.72)	27 (13.43)	176 (20.75)	151 (22.98)	25 (13.09)
Sex, n (%)						
Female	764 (90.95)	586 (91.71)	178 (88.56)	660 (77.83)	519 (79.00)	141 (73.82)
Male	71 (8.45)	48 (7.51)	23 (11.44)	186 (21.93)	136 (20.70)	50 (26.18)
Non-binary	5 (0.60)	5 (0.78)	0 (0.00)	2 (0.24)	2 (0.30)	0 (0.00)
Race, n (%)						
Asian	16 (1.90)	13 (2.03)	3 (1.49)	27 (3.18)	19 (2.89)	8 (4.19)
Black or African American	31 (3.69)	25 (3.91)	6 (2.99)	21 (2.48)	14 (2.13)	7 (3.66)
Multiple races	32 (3.81)	27 (4.23)	5 (2.49)	28 (3.30)	23 (3.50)	5 (2.62)
Alaska Native, American Indian, Native Hawaiian or other Pacific Islander	2 (0.24)	2 (0.31)	0 (0.00)	1 (0.12)	1 (0.15)	0 (0.0)
Other	6 (0.71)	4 (0.63)	2 (1.00)	8 (0.94)	7 (1.07)	1 (0.52)
White	753 (89.64)	568 (88.89)	185 (92.04)	763 (89.98)	593 (90.26)	170 (89.01)

[a]ILI: influenza-like illness.
[b]HTRI: Home Testing of Respiratory Illness.
[c]SD: Standard Deviatio.
[d]BMI: Body Mass Index

Table 3. Wearable sensor (activity) feature metrics during ILI[a] Days –4 to +1[b].

	FluStudy2020		HTRI[c]	
Wearable sensor (activity) metrics during ILI Days -4 to +1	**Influenza negative**	**Influenza positive**	**Influenza negative**	**Influenza positive**
HRV[d] features				
Modified CSI[e]	14111.93 (13550.93-14672.93)	16589.45 (15311.72-17867.19)	10814.64 (10450.41-11178.88)	11220.12 (10537.81-11902.43)
CSI	34.46 (33.77-35.16)	36.77 (35.57-37.98)	31.44 (30.75-32.12)	31.86 (30.69-33.04)
Cardiac vagal index	3.57 (3.54-3.60)	3.60 (3.55-3.64)	3.52 (3.50-3.54)	3.52 (3.48-3.56)
Variance in HRV in the high frequency	4.06 (3.82-4.30)	4.12 (3.70-4.54)	3.75 (3.55-3.95)	3.96 (3.51-4.40)
Normalized high frequency power	15.56 (15.52-15.60)	15.61 (15.55-15.67)	15.45 (15.41-15.48)	15.48 (15.41-15.54)
Variance in HRV in the low frequency	22.48 (21.07-23.89)	22.75 (20.30-25.21)	20.91 (19.75-22.06)	22.14 (19.48-24.79)
Low frequency / high frequency ratio	5.44 (5.42-5.45)	5.41 (5.39-5.44)	5.48 (5.47-5.50)	5.47 (5.44-5.50)
Normalized low frequency power	84.44 (84.40-84.48)	84.39 (84.33-84.45)	84.55 (84.52-84.59)	84.52 (84.46-84.59)
Total power density spectral	138.63 (129.90-147.35)	140.25 (125.06-155.43)	128.92 (121.78-136.06)	136.52 (120.10-152.94)
Variance in HRV in the very low frequency	112.09 (105.01-119.16)	113.37 (101.06-125.69)	104.27 (98.48-110.05)	110.43 (97.11-123.74)
HRV triangular index measurement	8.66 (8.49-8.83)	9.03 (8.75-9.30)	8.15 (8.00-8.29)	8.26 (8.00-8.53)
Percentage of minutes with HR[f] >1.5× RHR[g] for the day	0.06 (0.06-0.07)	0.07 (0.06-0.08)	0.07 (0.06-0.07)	0.08 (0.07-0.09)
Mean RHR during main sleep	71.61 (70.79-72.42)	72.87 (71.55-74.20)	69.43 (68.72-70.13)	70.45 (69.11-71.80)
Standard deviation of the projection of the Poincare plot on the line perpendicular to the line of identity	2.96 (2.85-3.06)	2.95 (2.78-3.12)	2.90 (2.81-2.99)	2.94 (2.76-3.12)
Standard deviation of the projection of the Poincare plot on the line of identity	95.86 (93.10-98.63)	102.11 (97.13-107.10)	83.58 (81.62-85.53)	84.49 (80.77-88.20)
Coefficient of variation equal to the ratio of SDNN[h] divided by mean nni[i]	0.08 (0.08-0.08)	0.09 (0.08-0.09)	0.07 (0.07-0.07)	0.07 (0.07-0.07)
Coefficient of variation	0.0047 (0.0046-	0.0048 (0.0046-	0.0045	0.0046

of successive differences equal to the RMSSD[j] divided by mean nni	0.0049)	0.0050)	(0.0044-0.0046)	(0.0044-0.0048)
Maximum HR	95.25 (94.24-96.26)	99.11 (97.52-100.69)	85.30 (84.55-86.05)	86.85 (85.41-88.29)
Mean HR	72.19 (71.38-73.00)	73.61 (72.30-74.93)	69.77 (69.07-70.47)	70.83 (69.48-72.17)
Mean of RR intervals[k] (mean nni)	857.35 (847.27-867.42)	843.26 (827.44-859.07)	882.15 (873.05-891.25)	873.42 (856.42-890.41)
Median absolute values of the successive differences between the RR intervals	863.51 (853.09-873.94)	850.66 (834.18-867.13)	884.26 (875.02-893.50)	876.11 (858.94-893.28)
Minimum HR	60.82 (60.08-61.55)	61.08 (59.93-62.24)	59.59 (58.95-60.22)	60.40 (59.16-61.64)
Number of interval differences of successive RR intervals >20 ms[l] (nni 20)	142.80 (137.48-148.12)	151.23 (142.69-159.77)	140.38 (135.67-145.10)	138.85 (129.86-147.84)
Number of interval differences of successive RR intervals >50 ms (nni 50)	44.78 (42.34-47.22)	48.09 (43.70-52.49)	42.87 (40.65-45.08)	43.37 (39.09-47.65)
Proportion derived by dividing nni 20 by the total number of RR intervals	0.52 (0.50-0.54)	0.51 (0.48-0.54)	0.52 (0.50-0.54)	0.52 (0.49-0.55)
Proportion derived by dividing nni 50 by the total number of RR intervals	0.17 (0.16-0.18)	0.17 (0.15-0.18)	0.16 (0.15-0.17)	0.17 (0.15-0.18)
Difference between the maximum and minimum nni	361.51 (352.28-370.73)	382.60 (366.46-398.74)	313.82 (306.69-320.95)	315.48 (302.11-328.85)
RMSSD	4.18 (4.04-4.32)	4.17 (3.93-4.41)	4.10 (3.97-4.23)	4.16 (3.90-4.42)
SDNN	67.82 (65.87-69.78)	72.24 (68.71-75.77)	59.14 (57.75-60.52)	59.78 (57.15-62.41)
SDSD[m]	4.18 (4.04-4.32)	4.17 (3.93-4.41)	4.10 (3.97-4.23)	4.16 (3.90-4.42)
Standard deviation of HR	6.13 (5.97-6.30)	6.86 (6.53-7.20)	4.74 (4.65-4.84)	4.93 (4.75-5.12)
Mean RHR	68.90 (68.16-69.64)	68.84 (67.61-70.08)	68.40 (67.74-69.05)	67.98 (66.80-69.15)
Sleep features				

Mean sleep duration	432.32 (423.96-440.67)	471.19 (458.19-484.19)	429.51 (422.04-436.99)	439.95 (424.71-455.18)
Mean sleep efficiency	88.42 (87.36-89.47)	90.76 (89.48-92.04)	88.35 (87.34-89.37)	88.02 (86.13-89.91)
Minutes in bed for the main sleep only of the day	457.60 (450.33-464.87)	484.73 (472.40-497.06)	462.63 (456.69-468.57)	461.90 (449.28-474.53)
Mean nap count	0.27 (0.25-0.29)	0.28 (0.24-0.32)	0.21 (0.19-0.23)	0.33 (0.29-0.37)
Mean total minutes in bed	493.68 (486.26-501.11)	526.61 (513.75-539.46)	490.56 (484.49-496.64)	506.42 (493.17-519.66)
Steps and activity features				
Proportion of the day the participant spent being physically active (≥50 steps per minute)	0.13 (0.13-0.14)	0.13 (0.12-0.14)	0.15 (0.14-0.15)	0.15 (0.14-0.16)
Proportion of minutes with non-zero steps out of the total minutes the device was worn	0.18 (0.17-0.18)	0.17 (0.16-0.18)	0.18 (0.18-0.19)	0.18 (0.17-0.19)
Maximum amount of activity the participant was able to complete within a single hour of the day	1812.75 (1719.70-1905.80)	1753.13 (1595.91-1910.35)	1930.59 (1825.67-2035.51)	1915.84 (1738.62-2093.07)
Total number of steps	7088.80 (6761.89-7415.71)	7100.59 (6590.66-7610.52)	7536.54 (7231.69-7841.39)	7487.53 (6906.04-8069.03)

[a]ILI: influenza-like illness.
[b]Values are presented as mean (95% confidence interval).
[c]HTRI: Home Testing of Respiratory Illness.
[d]HRV: heart rate variability.
[e]CSI: cardiac sympathetic index.
[f]HR: heart rate.
[g]RHR: resting heart rate.
[h]SDNN: mean of the standard deviations of normal-to-normal interval.
[i]nni: normal-to-normal interval.
[j]RMSSD: square root mean of the sum of the squares of differences between adjacent normal-to-normal intervals.
[k]RR interval: the time elapsed between two successive R-waves of the QRS signal on the electrocardiogram.
[l]ms: milliseconds.
[m]SDSD: standard deviation of successive differences.

Assessment of the XGBoost Model for Influenza Prediction During ILI Days –4 to +1

XGBoost models informed by symptoms-only data, activity-only data, or a combination of both symptoms and activity data were evaluated across training, validation, and test sets for FluStudy 2020. ROC curves and stratified k-fold cross-validation analyses for all models are presented in Figure 1, with confusion matrices shown in Figure S2 in Multimedia Appendix 1.

For the training and validation sets, the model trained on the combined symptoms and activity data (training AUC, 0.77; validation AUC, 0.74) consistently outperformed the models trained on the symptoms-only data (training AUC, 0.73; validation AUC, 0.72) and activity-only data (training AUC, 0.68; validation AUC, 0.65) (Figure 1). When applied to the FluStudy2020 test set, the model performance with the combined symptoms and activity data was closely aligned with the symptoms-only data (combined symptom and activity test AUC, 0.74; symptoms-only test AUC, 0.74). We extended our evaluation to the HTRI study, where the model trained on combined symptoms and activity data (evaluation AUC, 0.75) outperformed the model trained on the symptoms-only data (evaluation AUC, 0.74), confirming the results of the FluStudy2020 training and validation sets.

Feature importance plots for each model are presented in Figure 2. For the combined symptoms and activity model, cough, mean RHR during main sleep, fever, and total minutes in bed were the most important, with mean feature importance values of 0.21, 0.17, 0.15, and 0.15, respectively. For the symptoms-only model, fever, cough, and sore throat were the most important, with mean feature importance values of 0.36, 0.33, and 0.10, respectively. The heart low frequency / high frequency ratio, total minutes in bed, mean RHR during main sleep, and heart normalized low frequency power were the top features influencing activity-only model predictions, with mean feature importance values of 0.34, 0.17, 0.15, and 0.15, respectively (Figure 2). Calibration plots highlighting the degree of correspondence between the estimated probability of influenza-positive cases and observed influenza cases for each model are presented in Figure 3.

Figure 1. Receiver Operating Characteristic (ROC) curves for XGBoost model discrimination between influenza-positive and influenza-negative participants for FluStudy2020 and Home Testing of Respiratory Illness (HTRI) data. XGBoost model performance was assessed for

symptoms-only data, activity-only data, and a combination of symptoms and activity data. The mean performance across each k-fold and 95% confidence intervals (CI) for the training, validation, and test sets are presented. Mean values ± the margin of error are shown for: area under the curve (AUC), balanced accuracy (BA), sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), and weighted harmonic mean of precision and recall (F2). The red line represents random guess and the blue line represents mean ROC ± 95% CI.

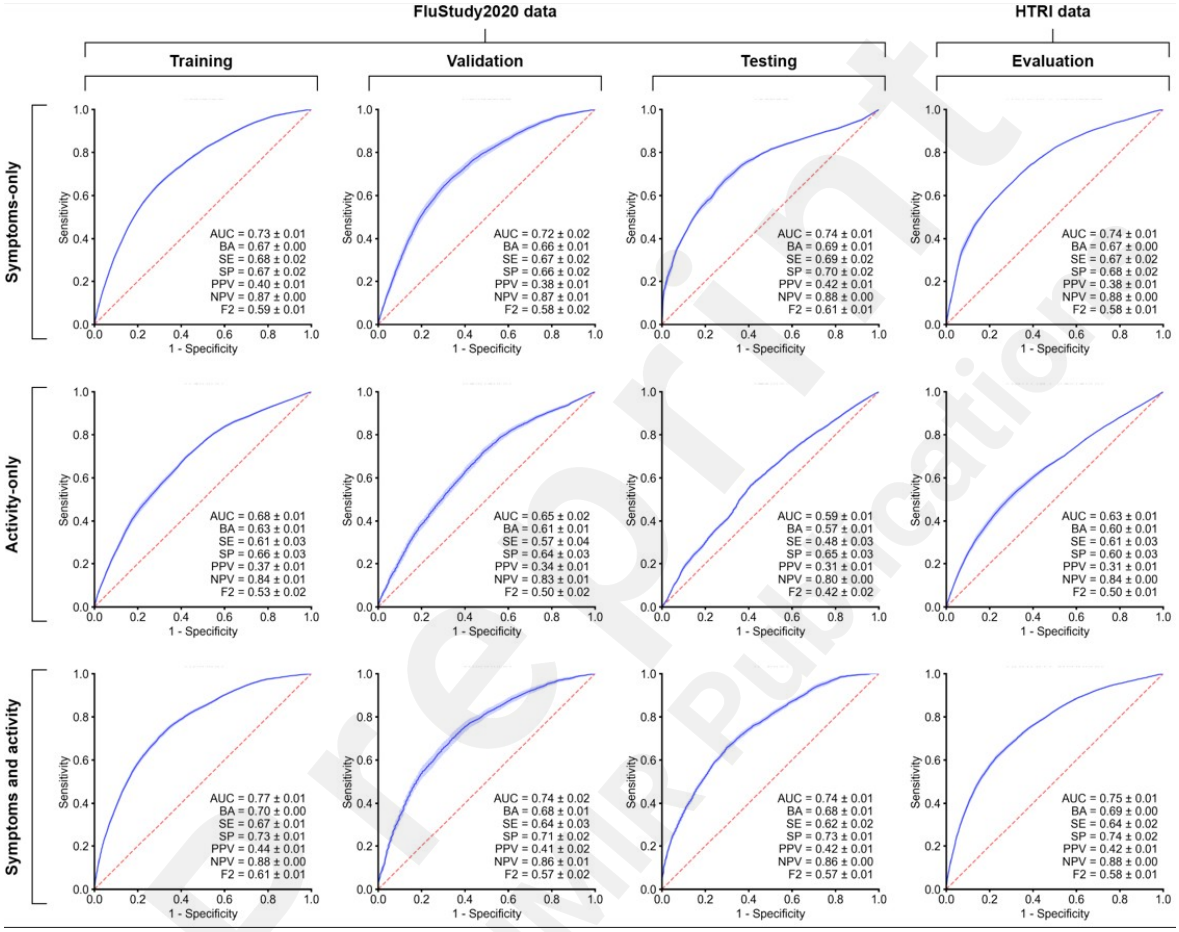


Figure 2. Feature importance plots for symptoms-only data (A), activity-only data (B), and a combination of symptoms and activity data (C). Values are presented as mean (95% confidence interval [CI]). BMI: body mass index; HR: heart rate; HRV: heart rate variability; RHR: resting heart rate; RR interval: the time elapsed between two successive R-waves of the QRS signal on the electrocardiogram.

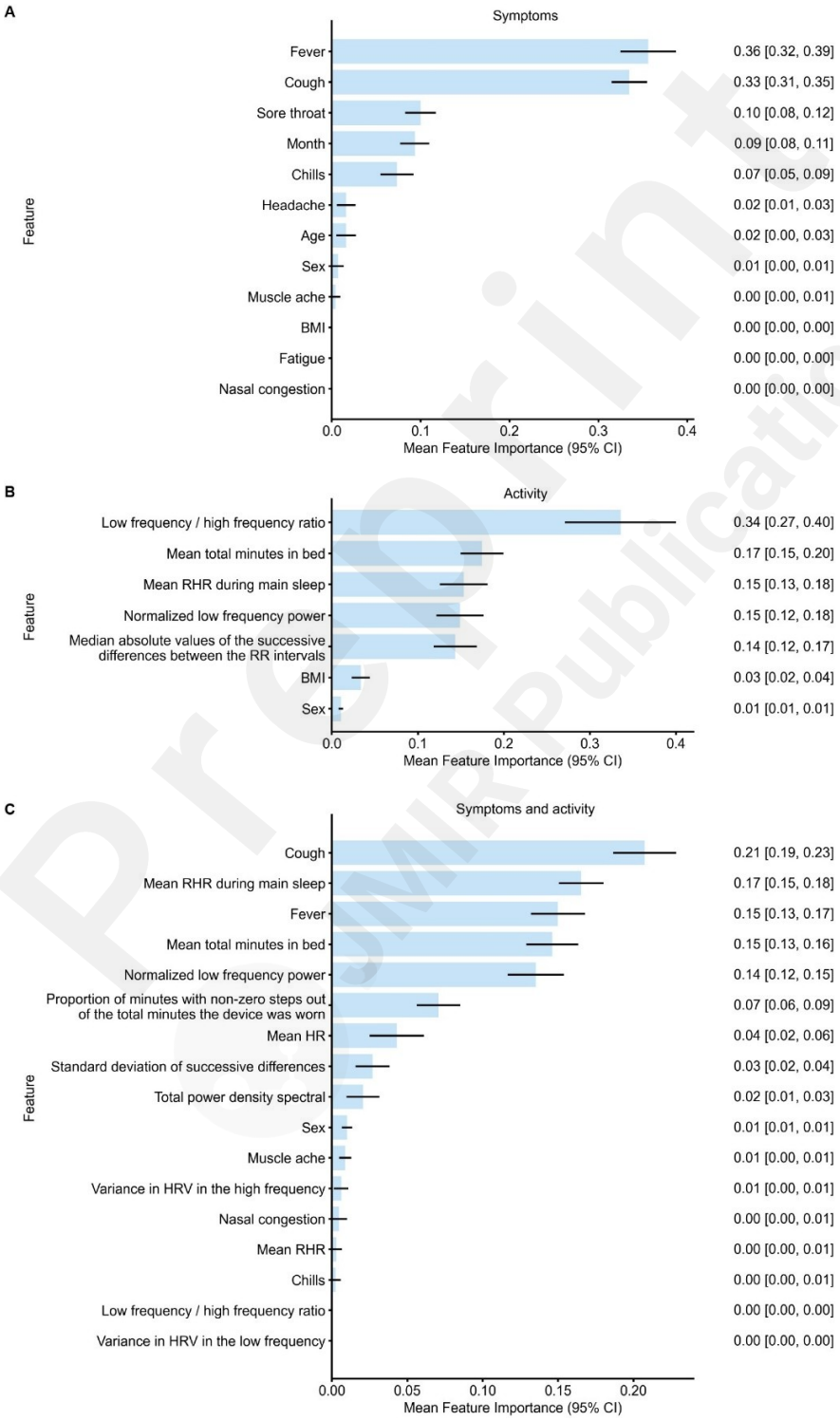
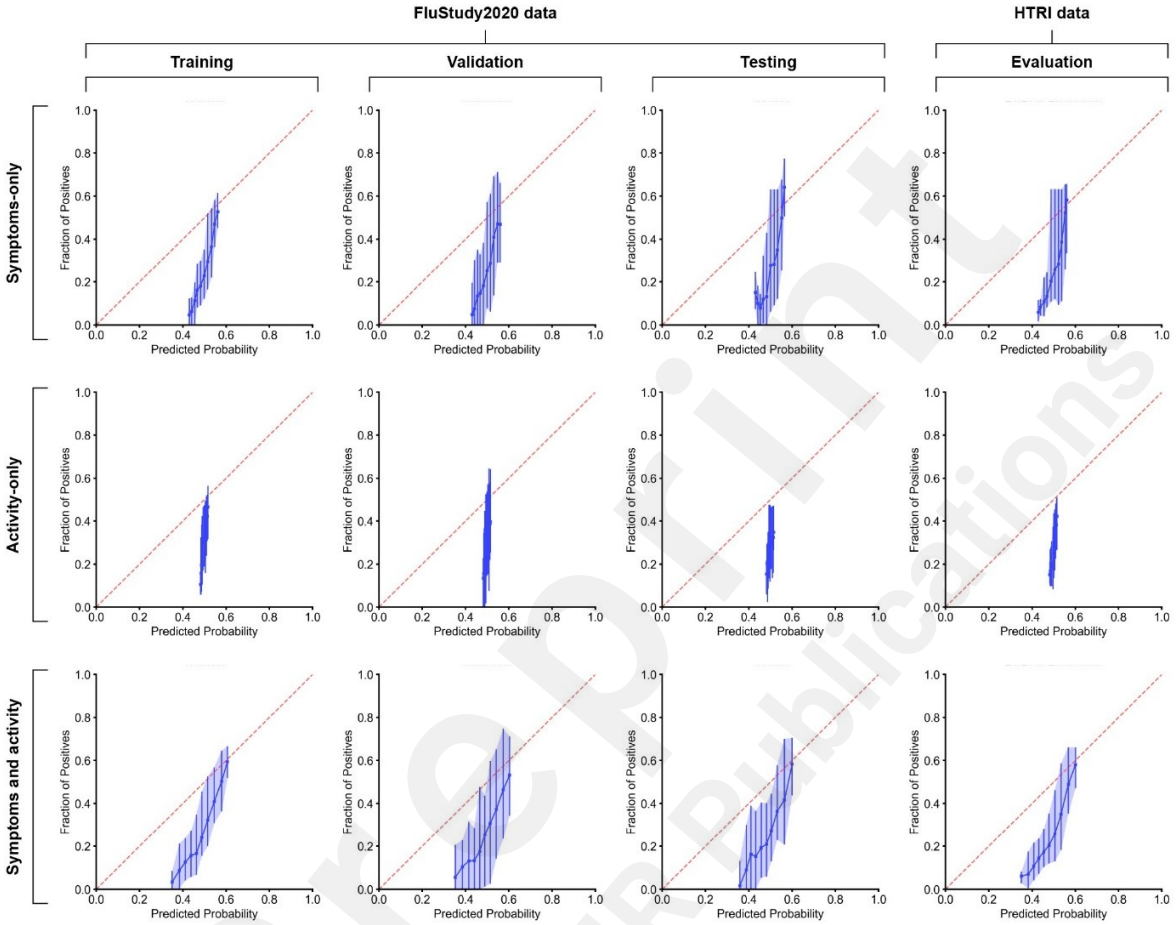


Figure 3. Calibration plots for FluStudy2020 and Home Testing of Respiratory Illness (HTRI) data, as assessed using symptoms-only data, activity-only data, and a combination of symptoms and activity data. The red dashed line represents perfect calibration and blue dots represent mean calibration ± 95% confidence interval (CI).



Discussion

To our knowledge, this is the largest study of commercial wearable sensors for the early detection of influenza incorporating virological confirmation of influenza infection. The study was specifically designed to test, in the real world, the hypothesis generated in experimental settings that wearable sensor data may predict the onset of viral respiratory infection. For the combined symptoms and activity model, the most important variables were cough, mean RHR during main sleep, fever, and total minutes in bed; for the symptoms-only model, the most important variables were fever, cough, and sore throat; and for the activity-only model, the most important variables were heart low frequency/high frequency ratio, total minutes in bed, mean RHR during main sleep, and heart normalized low frequency power. The best-performing machine learning model for influenza detection was trained on the combined symptoms and activity data, and had a mean training AUC of 0.77. Model performance was validated on an independent dataset (HTRI) not used for training, which yielded a mean AUC of 0.75. The accuracy of the combined machine learning model was further confirmed by calibration plots for combined symptoms and activity data, which were well calibrated compared with symptoms-only or activity-only plots. Our model performance is significantly lower than in other studies using machine-learning algorithms to predict influenza infection using wearable-sensor data, which achieved accuracies of up to 94% [4, 11]. However, these results were from small cohort (n=31 and n=20) challenge studies, where participants used research-grade wearable sensors and remained in controlled environments for up to 10 days post-challenge with either influenza A virus subtype H1N1 (A/H1N1) or influenza A virus subtype H3N2 (A/H3N2) [4, 11]. Grzesiak and colleagues noted that model accuracy was associated with both the knowledge of the timing and dosage of inoculation, and the high-fidelity measurements of research-grade sensors [4]. Our results suggest that findings from research-grade sensors tested in a highly controlled experimental setting may not easily translate to scalable low fidelity commercial-grade sensors deployed in the real world.

With the FluStudy2020 training and validation sets, the best-performing model for influenza detection used a combination of symptoms and activity features. However, in the FluStudy2020 test set, model performance was similar between the combined symptoms and activity data model and the symptoms-only data model. This implies the activity features do not significantly improve the model performance. In contrast, Quer and colleagues, using similar methods to discriminate between symptomatic individuals who were positive or negative for COVID-19, found that a model combining symptom and sensor data performed significantly better than one considering symptoms alone (AUC [IQR] 0.80 [0.73–0.86] vs 0.71 [0.63–0.79]) [5]. The different results observed may simply reflect differences between influenza and COVID-19, which have several non-overlapping symptoms; notably, their model included data from onset day to Day 7 whereas our model included data from Day –4 to Day 1. Our model restricted the data period as the clinical utility of wearable sensors as an early warning tool for influenza would depend on their ability to detect infection early in its course, when the individual could take action to limit spread or seek medical attention.

A key strength of our study is the laboratory confirmation of influenza in symptomatic patients using a highly accurate RT-PCR test, which provides accurate ascertainment of true positives and true negatives. Another strength of our study is that demographic, clinical, and Fitbit device data from a large, real-world population of more than 800 participants were used in model development. For the combined model (symptoms and activity data), cough, total minutes in bed, and mean RHR during main sleep were the top three features influencing

model predictions. Nine of the top 17 most important features influencing model predictions were HRV metrics. Deviations in HRV metrics have been associated with infection status and severity of various bacterial and viral illnesses [4, 7, 10-12, 22]. Additionally, Hirten et al. showed that SDNN (the mean amplitude of the circadian pattern of the SD of the inter-beat interval of normal sinus beats) was associated with a COVID-19 diagnosis, irrespective of symptomatology [22]. Another study demonstrated that HRV acrophase and HRV MESOR (circadian heart rate variability mean) were among the most important predictors of COVID-19 infection, along with age and BMI [8]. Future analyses should consider the impact of biological and lifestyle factors such as sex, menstrual cycle, and alcohol consumption on HRV and other physiological features [9, 10, 38, 39].

Limitations

Limitations pertaining to the HTRI and FluStudy2020 study design have been discussed previously [15]. Notably, the very small numbers of African Americans, Asians, males, and adults aged ≥65 years in this cohort limits the generalizability of the model. The imbalances may be the result of differences in the likelihood of these populations to engage with digital health services; for example, women have been found to be more likely to use a mobile health app than men [40]. In addition, participants were required to own a Fitbit device, which may predispose this cohort to exhibiting increased levels of activity and more health-conscious behaviors than the general population, which could limit the generalizability of the activity-based predictive models. A single device type (Fitbit) was used to minimize measurement error that could arise from the use of multiple device types in study participants. However, this limits the generalizability of the findings as several other device types are in widespread use. Symptom self-reporting used in both studies is subjective and prone to recall bias. However, results of influenza tests performed as part of the study were not provided to participants, which could have led to a differential recall of symptoms between influenza-positive and influenza-negative participants. Nevertheless, we cannot rule out participants' awareness of their disease status through seeking routine care for their ILI outside of the study. Additionally, symptoms data were collected daily to minimize the risk of incomplete or inaccurate recall.

The studies included in this analysis were designed prior to the COVID-19 pandemic but were ongoing until October 2020. COVID-19 mitigation measures such as lockdown procedures may have impacted participants' regular activities and influenza circulation during the period of these studies. Further implications of the COVID-19 pandemic have been discussed previously [15].

While our previous work demonstrates that the amplitude of wearable sensor deviations differs significantly between influenza-positive individuals and those with ILI symptoms only, the symptoms and activity features employed in model development in this study are not unique to influenza infection [15]. Additionally, strict symptom criteria were used to define the symptomatic ILI population, which may have selected a more severe symptomatic population and limited the ability to discriminate between influenza-positive and influenza-negative participants. Future studies with different study designs and less restrictive symptom eligibility criteria should investigate the ability of machine-learning algorithms to discriminate among a range of other common respiratory viral infections using symptomatic and wearable sensor data.

Finally, the validity and reliability of commercial wearable sensors in the measurement of steps, sleep, and heart rate have been the subject of debate. A systematic review including

over 150 publications found that FitBit heart rate measurements were variable and tended towards underestimating heart rate [41]. Additionally, the wearable devices used by participants in our study only measured steps, sleep, and heart rate. Future studies should investigate whether more advanced wearable sensors with more accurate accelerometers and including additional physiologic measures, such as skin temperature and blood oxygen saturation, could improve the performance of commercial-grade sensors in early detection and discrimination of respiratory viral infections.

Conclusions

We demonstrate that a machine-learning algorithm combining symptomatic and commercial wearable sensor data during the latent and early symptomatic phases of ILI had moderate accuracy to detect influenza in a large real-world cohort of symptomatic ILI individuals, suggesting that previous findings from research-grade sensors tested in highly controlled experimental settings may not easily translate to scalable commercial-grade sensors deployed in the real world. The model maintained consistent performance across two distinct studies. The model was initially trained and evaluated on FluStudy2020 data, and achieved comparable performance when validated on the HTRI data, affirming its generalizability. If machine learning algorithms using commercial wearable sensors had strong predictive power and were validated, they may potentially play a role in public health surveillance and could prompt users to adopt infection-control behavior (eg, self-quarantine) and to seek early medical attention if necessary. In the future, more advanced wearables measuring additional physiologic parameters may improve the performance of wearable sensors in the early detection and discrimination of viral respiratory infections.

Acknowledgments

The FluStudy2020 was supported by F. Hoffmann-La Roche Ltd, Genentech, Inc., and Evidation Health Inc. The HTRI study was supported by BARDA (Contract Number 75A50119C00036), part of the Office of the Assistant Secretary for Preparedness and Response, US Department of Health and Human Services, and Audere, a digital health, non-profit organization. The HTRI study was designed and conducted by Evidation Health Inc. F. Hoffmann-La Roche Ltd were involved in the design and conduct of FluStudy2020; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication.

These data were contributed by participants as part of the HTRI study developed by Evidation Health, Inc. and described in Synapse [ID: syn22803188]. Support with acquisition, analysis, and interpretation of the data was provided by Konrad Mysliwiec, Roche Global IT Solution Centre, Warsaw, Poland. Support with obtaining funding was provided by Kamran Farooq, Marco Prunotto, Ahmed Ansari, F. Hoffmann-La Roche Ltd, Basel, Switzerland, and James Harper, Roche Products Ltd, Welwyn Garden City, UK. Third-party medical writing assistance, under the direction of the authors, was provided by Stephanie Cumberworth, PhD, and Edmund Harratt, BA, on behalf of Ashfield MedComms, an Inizio company, and funded by F. Hoffmann-La Roche Ltd.

Author Contributions

ML and KF had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. KF, ML, AH-R, AO, LD-H, FJ, VU, DC, VH, MP, and BC conceived and designed the study. KF, ML, AH-R, AO, LD-H, FJ, KN, KM, MMAZ, DC, VH, BC, and VU conducted the acquisition, analysis, or interpretation of data.

LD-H, ML, AH-R, AO, and FJ conducted statistical analysis. KF, MP, AA, and JH obtained funding for the study. KF and ML provided administrative, technical, or material support. ML, KF, and VU supervised the writing of the manuscript. All authors contributed to the drafting of the manuscript and conducted critical revision for important intellectual content.

Conflicts of Interest

ML, VH, and DC are current or former employees of Genentech, Inc., a member of the Roche Group. LD-H, FJ, BC, and VU are current or former employees of Roche Products Ltd. KF, AO, AH-R, and MP are employees of F. Hoffmann-La Roche Ltd. MMAZ is an employee of Roche Services (Asia Pacific) Sdn. Bhd. KN is an employee of Badger Software Sp. z o.o. Badger Software Sp. z o.o. received funding from F. Hoffmann-La Roche Ltd for the conduct of this study but were not paid for the development of the manuscript.

Data Availability

The data from the Home Testing of Respiratory Illness study are available on the internet [42]. The data sets analyzed during this study are available from the corresponding author on reasonable request.

Abbreviations

AUC: area under curve
A/H1N1: influenza A virus subtype H1N1
A/H3N2: influenza A virus subtype H3N2
BA: balanced accuracy
BMI: body mass index
CI: confidence interval
CSI: circadian sympathetic index
F2: weighted harmonic mean of precision and recall
HR: heart rate
HRV: heart rate variability
HRV MESOR: circadian heart rate variability mean
HTRI: Home Testing of Respiratory Illness
ILI: influenza-like illness
IQR: interquartile range
ms: milliseconds
nni: normal-to-normal interval
NPV: negative predictive value
PPV: positive predictive value
RHR: resting heart rate
RMSSD: square root mean of the sum of the squares of differences between adjacent normal-to-normal intervals
ROC: receiver operating characteristics
RR interval: the time elapsed between two successive R-waves of the QRS signal on the electrocardiogram.
RT-PCR: reverse transcription polymerase chain reaction
SD: standard deviation
SDSD: standard deviation of successive differences
SDNN: mean amplitude of the circadian pattern of the standard deviation of the inter-beat interval of normal sinus beats
SE: sensitivity
SP: specificity
US: United States
XGBoost: extreme gradient boosting

Multimedia Appendix 1: Legends

Table S1. FluStudy2020 and HTRI inclusion and exclusion criteria.

Figure S1. Model evaluation schematic. AUC: area under the curve; F2: weighted harmonic mean of precision and recall; HTRI: Home Testing of Respiratory Illness; ROC: receiver operating characteristics.

Figure S2. Confusion matrices for XGBoost model discrimination between influenza-positive and influenza-negative participants. XGBoost model performance was assessed for symptoms-only data, activity-only data, and a combination of symptoms and activity data. Mean values ± the margin of error are presented. HTRI: Home Testing of Respiratory Illness.

References

1. United States Centers for Disease Control and Prevention. Disease burden of flu. 2022. https://www.cdc.gov/flu/about/burden/index.html [accessed 02 Nov 2022].
2. D'Haese PF, Finomore V, Lesnik D, Kornhauser L, Schaefer T, Konrad PE, et al. Prediction of viral symptoms using wearable technology and artificial intelligence: A pilot study in healthcare workers. PLoS One. 2021;16(10):e0257997. PMID: 34648513. doi: 10.1371/journal.pone.0257997.
3. Gadaleta M, Radin JM, Baca-Motes K, Ramos E, Kheterpal V, Topol EJ, et al. Passive detection of COVID-19 with wearable sensors and explainable machine learning algorithms. NPJ Digit Med. 2021 Dec 8;4(1):166. PMID: 34880366. doi: 10.1038/s41746-021-00533-1.
4. Grzesiak E, Bent B, McClain MT, Woods CW, Tsalik EL, Nicholson BP, et al. Assessment of the feasibility of using noninvasive wearable biometric monitoring sensors to detect influenza and the common cold before symptom onset. JAMA Netw Open. 2021 Sep 1;4(9):e2128534. PMID: 34586364. doi: 10.1001/jamanetworkopen.2021.28534.
5. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med. 2021 Jan;27(1):73-7. PMID: 33122860. doi: 10.1038/s41591-020-1123-x.
6. Group A-TL-CS, Lundgren JD, Grund B, Barkauskas CE, Holland TL, Gottlieb RL, et al. A neutralizing monoclonal antibody for hospitalized patients with COVID-19. N Engl J Med. 2021 Mar 11;384(10):905-14. PMID: 33356051. doi: 10.1056/NEJMoa2033130.
7. Ahmad S, Tejuja A, Newman KD, Zarychanski R, Seely AJ. Clinical review: a review and analysis of heart rate variability and the diagnosis and prognosis of infection. Crit Care. 2009;13(6):232. PMID: 20017889. doi: 10.1186/cc8132.
8. Hirten RP, Tomalin L, Danieletto M, Golden E, Zweig M, Kaur S, et al. Evaluation of a machine learning approach utilizing wearable data for prediction of SARS-CoV-2 infection in healthcare workers. JAMIA Open. 2022 Jul;5(2):ooac041. PMID: 35677186. doi: 10.1093/jamiaopen/ooac041.
9. Mitratza M, Goodale BM, Shagadatova A, Kovacevic V, van de Wijgert J, Brakenhoff TB, et al. The performance of wearable sensors in the detection of SARS-CoV-2 infection: a systematic review. Lancet Digit Health. 2022 May;4(5):e370-e83. PMID: 35461692. doi: 10.1016/S2589-7500(22)00019-X.
10. Radin JM, Quer G, Pandit JA, Gadaleta M, Baca-Motes K, Ramos E, et al. Sensor-based surveillance for digitising real-time COVID-19 tracking in the USA (DETECT): a multivariable, population-based, modelling study. Lancet Digit Health. 2022 Nov;4(11):e777-e86. PMID: 36154810. doi: 10.1016/S2589-7500(22)00156-X.
11. Temple DS, Hegarty-Craver M, Furberg RD, Preble EA, Bergstrom E, Gardener Z, et al. Wearable sensor-based detection of influenza in presymptomatic and asymptomatic individuals. J Infect Dis. 2022 Jun 27. PMID: 35759279. doi: 10.1093/infdis/jiac262.
12. Shandhi MMH, Cho PJ, Roghanizad AR, Singh K, Wang W, Enache OM, et al. A method for intelligent allocation of diagnostic testing by leveraging data from commercial wearable devices: a case study on COVID-19. NPJ Digit Med. 2022 Sep 1;5(1):130. PMID: 36050372. doi: 10.1038/s41746-022-00672-z.
13. Dhingra LS, Aminorroaya A, Oikonomou EK, Nargesi AA, Wilson FP, Krumholz HM, et al. Use of wearable devices in individuals with or at risk for cardiovascular disease in the US, 2019 to 2020. JAMA Netw Open. 2023 Jun 1;6(6):e2316634. PMID: 37285157. doi: 10.1001/jamanetworkopen.2023.16634.

14. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res. 2020 Oct 16;22(10):e22443. PMID: 33064083. doi: 10.2196/22443.
15. Hunter V, Shapiro A, Chawla D, Drawnel F, Ramirez E, Phillips E, et al. Characterization of influenza-like illness burden using commercial wearable sensor data and patient-reported outcomes: mixed methods cohort study. J Med Internet Res. 2023 Mar 23;25:e41050. PMID: 36951890. doi: 10.2196/41050.
16. Chen T, Guestrin C, editors. A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; New York, NY, USA: ACM.
17. developers X. XGBoost documentation. 2022. https://xgboost.readthedocs.io/en/stable/ [accessed 02 Nov 2022].
18. Mezlini A, Shapiro A, Daza EJ, Caddigan E, Ramirez E, Althoff T, et al. Estimating the burden of influenza-like illness on daily activity at the population scale using commercial wearable sensors. JAMA Netw Open. 2022 May 2;5(5):e2211958. PMID: 35552722. doi: 10.1001/jamanetworkopen.2022.11958.
19. Radin JM, Wineinger NE, Topol EJ, Steinhubl SR. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. Lancet Digit Health. 2020 Feb;2(2):e85-e93. PMID: 33334565. doi: 10.1016/S2589-7500(19)30222-5.
20. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng. 2020 Dec;4(12):1208-20. PMID: 33208926. doi: 10.1038/s41551-020-00640-6.
21. Natarajan A, Su HW, Heneghan C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. NPJ Digit Med. 2020 Nov 30;3(1):156. PMID: 33299095. doi: 10.1038/s41746-020-00363-7.
22. Hirten RP, Danieletto M, Tomalin L, Choi KH, Zweig M, Golden E, et al. Use of physiological data from a wearable device to identify SARS-CoV-2 infection and symptoms and predict COVID-19 diagnosis: observational study. J Med Internet Res. 2021 Feb 22;23(2):e26107. PMID: 33529156. doi: 10.2196/26107.
23. AH C, AK T, F G, C D. Wearable devices for the detection of COVID-19. Nat Electron. 2021;4:13-4. doi: 10.1038/s41928-020-00533-1.
24. Cleary JL, Fang Y, Sen S, Wu Z. A caveat to using wearable sensor data for COVID-19 detection: The role of behavioral change after receipt of test results. PLoS One. 2022;17(12):e0277350. PMID: 36584148. doi: 10.1371/journal.pone.0277350.
25. Shapiro A, Marinsek N, Clay I, Bradshaw B, Ramirez E, Min J, et al. Characterizing COVID-19 and influenza illnesses in the real world via person-generated health data. Patterns (N Y). 2021 Jan 8;2(1):100188. PMID: 33506230. doi: 10.1016/j.patter.2020.100188.
26. Mekhael M, Lim CH, El Hajjar AH, Noujaim C, Pottle C, Makan N, et al. Studying the effect of long COVID-19 infection on sleep quality using wearable health devices: observational study. J Med Internet Res. 2022 Jul 5;24(7):e38000. PMID: 35731968. doi: 10.2196/38000.
27. Kotnik JH, Cooper S, Smedinghoff S, Gade P, Scherer K, Maier M, et al. Flu@home: the comparative accuracy of an at-home influenza rapid diagnostic test using a prepositioned test kit, mobile app, mail-in reference sample, and symptom-based testing trigger. J Clin Microbiol. 2022 Mar 16;60(3):e0207021. PMID: 35107302. doi: 10.1128/JCM.02070-21.
28. Saint-Maurice PF, Troiano RP, Bassett DR, Jr., Graubard BI, Carlson SA, Shiroma EJ,

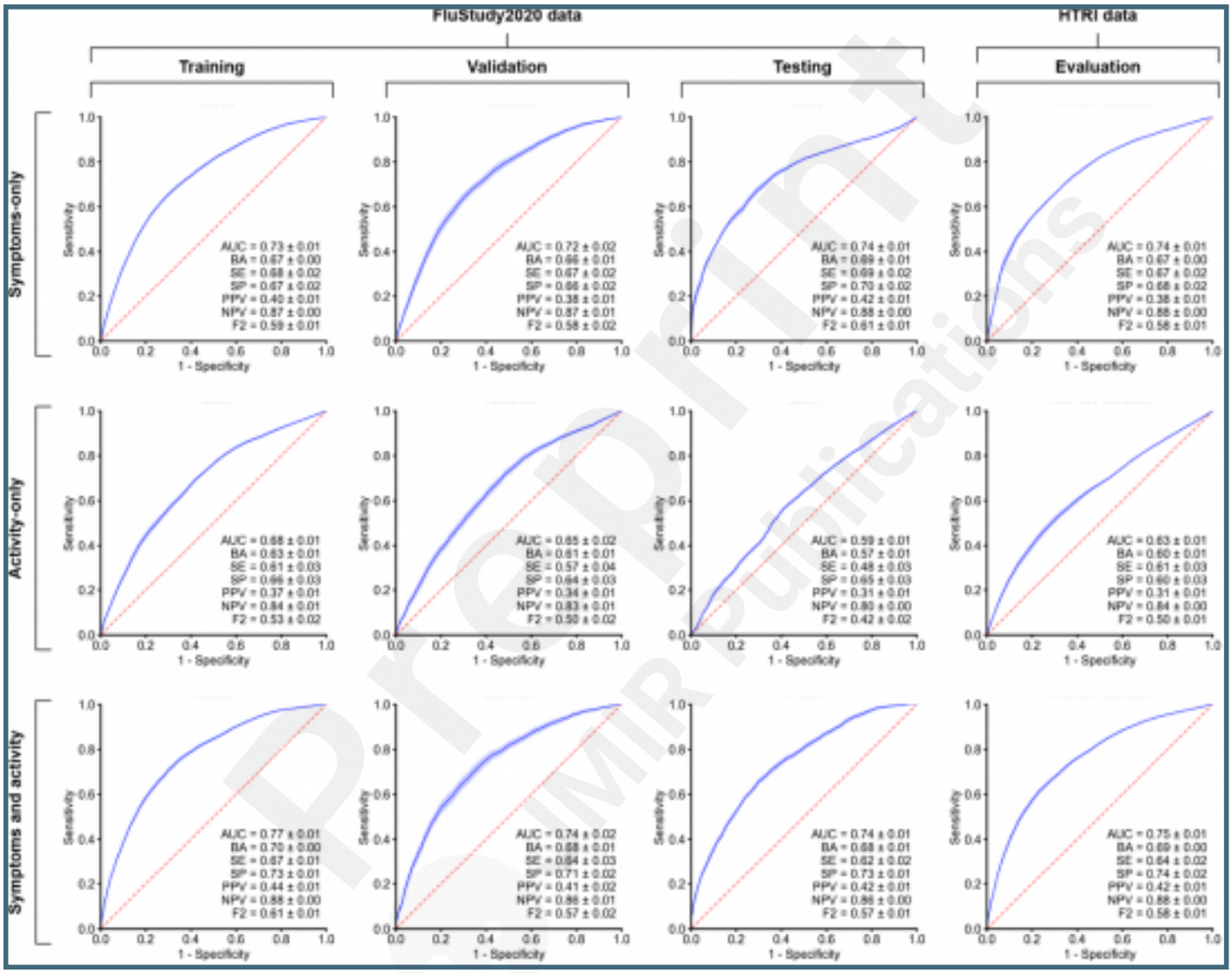
et al. Association of daily step count and step intensity with mortality among US adults. JAMA. 2020 Mar 24;323(12):1151-60. PMID: 32207799. doi: 10.1001/jama.2020.1382.

29. Master H, Annis J, Huang S, Beckman JA, Ratsimbazafy F, Marginean K, et al. Association of step counts over time with the risk of chronic disease in the All of Us Research Program. Nat Med. 2022 Nov;28(11):2301-8. PMID: 36216933. doi: 10.1038/s41591-022-02012-w.
30. McKinney W. Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference. 2010:56-61. doi: 10.25080/Majora-92bf1922-00a.
31. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020 Sep;585(7825):357-62. PMID: 32939066. doi: 10.1038/s41586-020-2649-2.
32. Hunter J. Matplotlib: A 2D Graphics Environment. Comput Sci Eng. 2007;9(3):90-5. doi: 10.1109/MCSE.2007.55.
33. Alam S, Chan N, Couto L, Dada Y, Danov I, Datta D, et al. Kedro (Version 0.18.12) [Computer software]. 2023. https://github.com/kedro-org/kedro [accessed 28 Sept 2023].
34. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw. 2010;36(11):1-13. doi: 10.18637/jss.v036.i11.
35. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020 Mar;17(3):261-72. PMID: 32015543. doi: 10.1038/s41592-019-0686-2.
36. Pedregosa F, Varoquaux G, Gramfort A, Vincent M, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-30. doi: 10.48550/arXiv.1201.0490.
37. Champseix R, Ribiere L, Le Couedic C. A Python package for heart rate variability analysis and signal preprocessing. J Open Res Softw. 2021;9(1):28. doi: 10.5334/jors.305.
38. Goodale BM, Shilaih M, Falco L, Dammeier F, Hamvas G, Leeners B. Wearable sensors reveal menses-driven changes in physiology and enable prediction of the fertile window: observational study. J Med Internet Res. 2019 Apr 18;21(4):e13404. PMID: 30998226. doi: 10.2196/13404.
39. Zhu G, Li J, Meng Z, Yu Y, Li Y, Tang X, et al. Learning from large-scale wearable device data for predicting the epidemic trend of COVID-19. Discrete Dyn Nat Soc. 2020 2020/05/05;2020:6152041. doi: 10.1155/2020/6152041.
40. Montagni I, Cariou T, Feuillet T, Langlois E, Tzourio C. Exploring digital health use and opinions of university students: field survey study. JMIR Mhealth Uhealth. 2018 Mar 15;6(3):e65. PMID: 29549071. doi: 10.2196/mhealth.9131.
41. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth. 2020 Sep 8;8(9):e18694. PMID: 32897239. doi: 10.2196/18694.
42. Evidation. Home testing of respiratory illness study research portal. 2020. https://www.synapse.org/#!Synapse:syn22803188/wiki/606343 [accessed 27 Sept 2023].

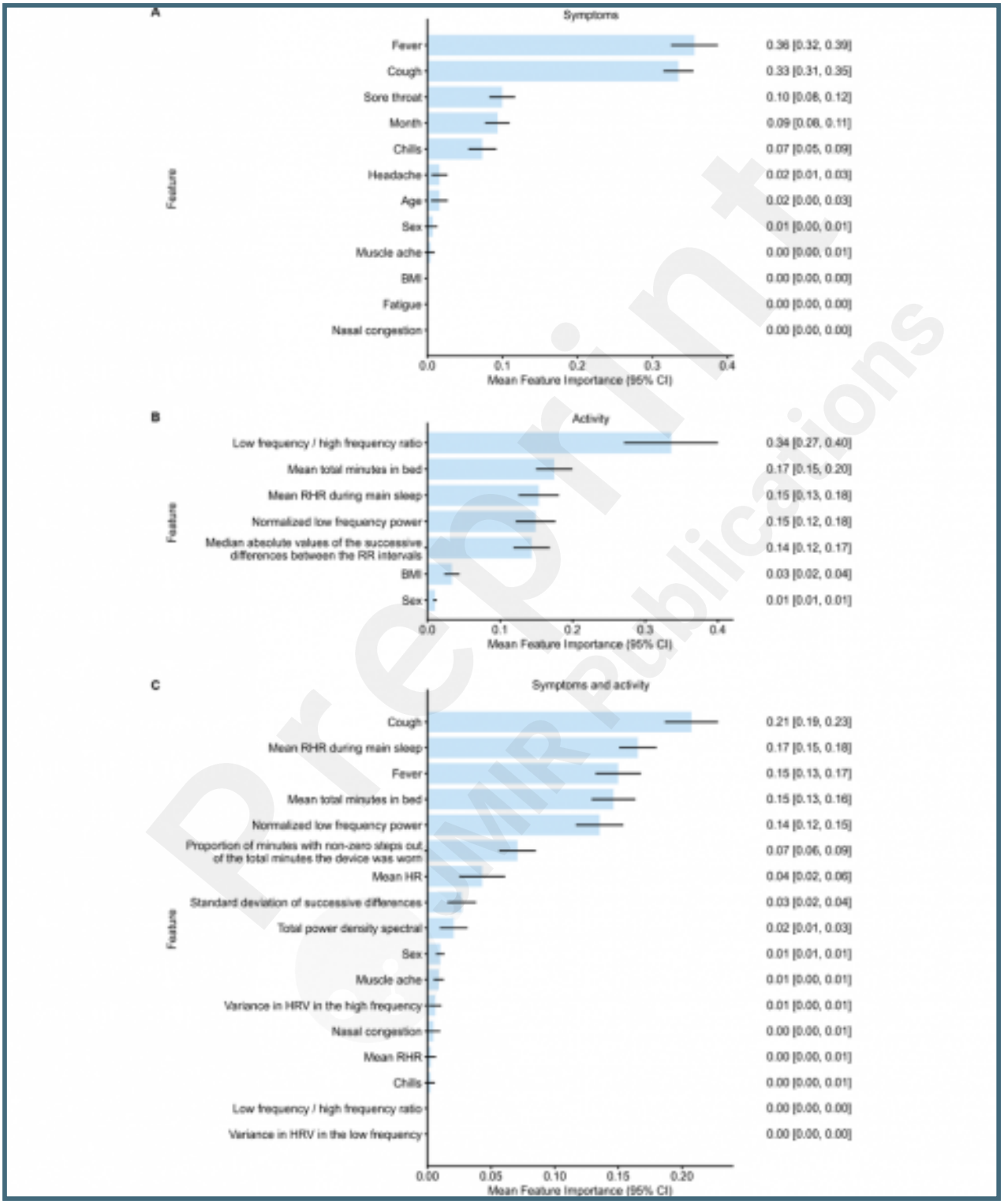
Supplementary Files

Figures

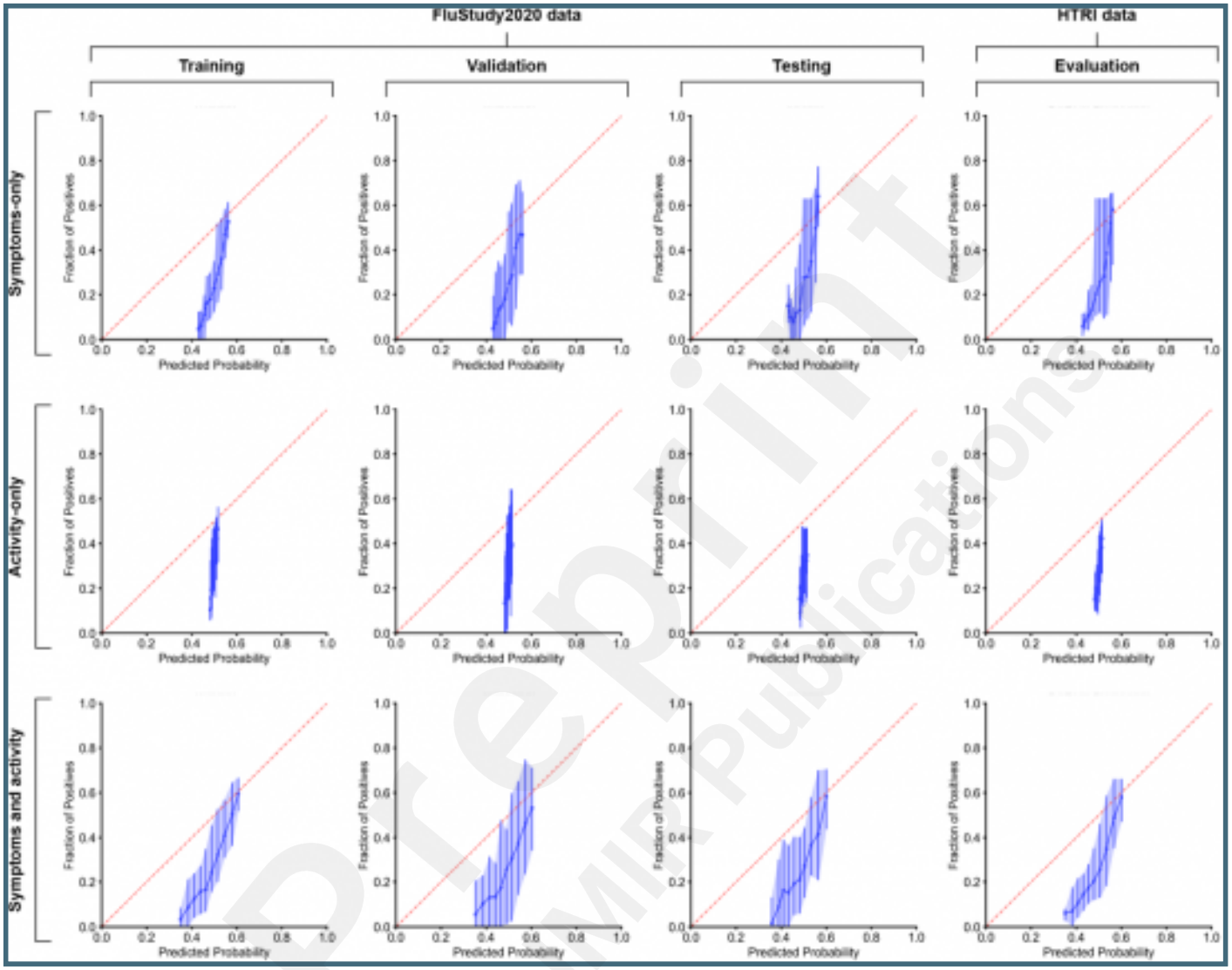
Receiver Operating Characteristic (ROC) curves for XGBoost model discrimination between influenza-positive and influenza-negative participants for FluStudy2020 and Home Testing of Respiratory Illness (HTRI) data. XGBoost model performance was assessed for symptoms-only data, activity-only data, and a combination of symptoms and activity data. The mean performance across each k-fold and 95% confidence intervals (CI) for the training, validation, and test sets are presented. Mean values ± the margin of error are shown for: area under the curve (AUC), balanced accuracy (BA), sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), and weighted harmonic mean of precision and recall (F2). The red line represents random guess and the blue line represents mean ROC ± 95% CI.



Feature importance plots for symptoms-only data (A), activity-only data (B), and a combination of symptoms and activity data (C). Values are presented as mean (95% confidence interval [CI]). BMI: body mass index; HR: heart rate; HRV: heart rate variability; RHR: resting heart rate; RR interval: the time elapsed between two successive R-waves of the QRS signal on the electrocardiogram.



Calibration plots for FluStudy2020 and Home Testing of Respiratory Illness (HTRI) data, as assessed using symptoms-only data, activity-only data, and a combination of symptoms and activity data. The red dashed line represents perfect calibration and blue dots represent mean calibration ± 95% confidence interval (CI).



Multimedia Appendixes

Untitled.
URL: http://asset.jmir.pub/assets/501ebae20320c6a6a035a186772d3f3a.pdf