# COVID Prognosticator: COVID-19 Onset With Wearable Devices

CHARLES J. CARVER, Dartmouth College, USA
HO MAN COLMAN LEUNG, Dartmouth College, USA
VINEY REGUNATH, Dartmouth College, USA
HUNTER GALLANT, Dartmouth College, USA

Over the past two years, the COVID-19 virus has significantly impacted the entire world and disrupted the global status quo. Unlike previous illnesses, COVID-19 is unique in its highly-transmissible and equally dangerous nature, infecting and killing of people. To combat COVID-19, early detection and intervention is essential, thereby limiting the spread and reach of the virus. Wearable devices, with their plethora of sensors and high-resolution data, offer a unique opportunity for predicting the onset of COVID-19 before it has the chance to spread. Prior work has already investigated the feasibility of using heart rate variability to detect the onset of COVID-19 with sufficient accuracy. This study builds off this idea, verifying that heart rate variability is a meaningful feature in COVID-19 onset detection using a publicly available dataset. We then expand on this claim by introducing data from additional sensors, including step count, pulse rate, and caloric data, and analyzing the performance of various machine learning models at predicting COVID-19 onset.

Additional Key Words and Phrases: Wearable devices, COVID-19

**ACM Reference Format:**
Charles J. Carver, Ho Man Colman Leung, Viney Regunath, and Hunter Gallant. 2022. COVID Prognosticator: COVID-19 Onset With Wearable Devices. 1, 1 (May 2022), 12 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

Given the continued relentlessness of COVID-19, proactive detection is essential for long-term management of the disease. Even after the original and delta variant outbreaks, nations continue to grapple with dealing with the virus. For instance, despite China's successful zero COVID-19 policy in the previous year, incidents of city wide lock downs have shown that there needs to be changes in policy of how governments manage COVID-19 cases [3]. One of the biggest problems surrounding the virus has been effectively detecting infected individuals. In 2020, several nations developed and utilized COVID-19 notification smartphone apps to attempt to control viral waves but those have led to unsatisfactory adoption and effectiveness. Furthermore, privacy and security concerns are well documented surrounding these applications [1]. In 2022, wearable devices appear offer an elegant solution to these concerns, given their ubiquity and rich set of collected data.

Prior work has already shown the possibility of using heart rate variability to detect the onset of COVID-19 using wearable device data [5, 7, 11, 14]. However, little work has been conducted on exploring how other sensing features (e.g., sleep duration, sleep variability) fare at predicting the onset of COVID-19, and even less on detecting the recovery of COVID-19. Jahrami et. al 2021 mentions how there is a correlation between higher rates of sleep problems and COVID-19 patients [6]. However, many studies such as Boscolo et al. mentioned that it is difficult to diagnose patients with COVID-19 without the use of antigen or PCR tests due to how similar said symptoms are to those of the common flu [2]. Furthermore, Boscolo mentions that the prevalence of COVID-19 among those with flu-like symptoms can vary substantially based on geographical context and phase of the disease. Studies from Tostmann et al. and Zayet et al [15, 16]. confirm this problem. The Wellroy dataset [9] provides us an unprecedented opportunity to analyze COVID-19 trends throughout the world. This project aims to explore the additional features provided by our dataset [9] at both predicting COVID19 onset and recovery. We hope

Authors' addresses: Charles J. Carver, ccarver.gr@dartmouth.edu, Dartmouth College, 15 Thayer Driver, Hanover, New Hampshire, USA, 03766; Ho Man Colman Leung, Dartmouth College, 15 Thayer Driver, Hanover, New Hampshire, USA, 03766; Viney Regunath, Dartmouth College, 15 Thayer Driver, Hanover, New Hampshire, USA, 03766; Hunter Gallant, Dartmouth College, 15 Thayer Driver, Hanover, New Hampshire, USA, 03766.

to infer new trends based on new studies on the disease and also investigate where wearable technology can contribute greatly to the field of epidemiology.

### 1.1 Project Objectives and Research Plan

This project aims to explore the additional features provided by a publicly available dataset [9] at both predicting COVID-19 onset and recovery. Features include heart rate data (e.g., time series data, variability), exercise information (e.g., steps taken, gait speed, calories burned), sleep data (e.g., sleep duration, sleep variability), and clinically-validated health surveys (e.g., COVID-19 symptoms, additional health symptoms). Instead of focusing solely on heart-rate variability, we will analyze these other features and determine their relevance at predicting COVID-19 offset and recovery.

To accomplish these goals, our approach is four-pronged: first, we will identify relevant health-factors from the existing body of literature that are promising indicators of COVID-19 onset; second, we will pre-process our dataset to ensure the features can be utilized in machine learning models; third, we will analyze the performance of multiple machine learning models on predicting the onset of COVID-19 using our pre-processed data; fourth, we will re-utilize these features and models to predict the recovery of COVID-19.

### 1.2 Contributions

In this paper, we evaluate the performance of 6 machine learning models in predicting COVID-19 onset using both heart rate variabtiliy data and wearable data retrieved from a publicly available dataset. We integrate data from multiple modalities to generate powerful features, enabling us to overcome the large variations in wearable data. Finally, we utilize a deep recurrent neural network to extract temporal features and identify correlations between COVID-19 onset with heart rate variability data and wearable data.

Our main contributions can be summarized as follows:

- Evaluate the correlation between COVID-19 onset, heart rate variability data, and wearable data using 6 machine learning models as a baseline validation.
- Utilize a deep recurrent neural network that includes wearable features to predict the onset of COVID-19.

## 2 RELATED

As discussed previously in Section 1, heart rate variability has been studied in the context of COVID-19, and continues to be a popular topic for research [5, 7, 11, 14]. Articles reviewing these studies show they can be markedly successful, with some of the results showing over eighty percent accuracy in predicting positive COVID-19 diagnoses [4, 13].

Many of these papers focused on traditional machine learning methods [7, 11]. Papers using these models had to perform a lot of auxiliary data processing, which, when combined with the standard preprocessing of health datasets, meant that a lot of time was spent getting the data ready instead of testing their predictive models. Research that focused on neural networks [5] was able to be completed without as much additional data processing. While the specific focii of these papers are different from our own, they serve as good references for how other researchers have built models that connect wearable data to COVID-19 symptoms.

In addition to the previously outlined methods, there has also been some evidence that certain health metrics change during periods of sleep when people become infected with COVID-19. During sleep, people naturally experience a difference in heart and respiration rate, but those who test positive for COVID-19 have greater differences in those metrics [8]. Similar to the studies focused primarily on heart rate variability, there are also correlations that could predict COVID-19 severity with high levels of accuracy. While this incorporates respiration rate in addition to heart rate, there are still more metrics that could be used to train machine learning models,

Table 1. Demographic information for COVID-19 dataset.

| Gender | | | Age | | | Geographic Location | | |
|---|---|---|---|---|---|---|---|---|
| | N | Percent | | N | Percent | | N | Percent |
| Male | 67 | 36% | 18-34 | 62 | 33% | Russia | 65 | 35% |
| Female | 118 | 64% | 35-44 | 56 | 30% | USA | 59 | 32% |
| | | | 45-74 | 68 | 37% | 23 Others | 61 | 33% |

such as movement and activity data from wearable technology, potentially further improving the prediction accuracy of COVID-19.

## 3 METHODS

### 3.1 Data Description

This study leverages the publicly available Welltory COVID-19 and Wearables Open Data Research dataset [9]. The dataset was collected during the Welltory's 2020 research into COVID-19. During the study, users with clinically validated COVID-19 statuses had their heart rate variability, symptoms, and wearable device data recorded using a mobile application. In terms of heart rate variability, the dataset contained signals recorded from any "Bluetooth-enabled heart rate monitor" or using Photoplethysmography (PPG) using a smartphone camera [9]. The wearable device data came from a variety of sources, including Apple Watches and Garmin devices. Finally, the dataset also includes self-reported symptom surveys from the individuals enrolled in the study.

As shown in Table 1, there were 185 participants, 36.2% identifying as male with the remaining 63.8% identifying as female. Over 30% of participants were between the ages of 35–44, roughly 33% between 18–34, and the remaining 37% between 45–74. In terms of location, over 35% of participants lived in Russia, over 32% in the United States, and the remaining third spread across 23 other countries.

The self-reported health surveys are composed of 74 distinct categories, including data on cholesterol, diabetes, and underlying health conditions (e.g., kidney, liver, and lung health). There are 11 questions specific to COVID-19. Out of these questions, 8 relate to symptom intensity for various COVID-19 symptoms (e.g., coughing, shortness of breath, fatigue). The remaining 3 questions are overall assessments of the COVID-19 diagnosis: "How long the user has been experiencing symptoms," "Overall state," and "An assessment of the user's symptoms which shows how likely it is that the user has coronavirus." Of the COVID-19 questions, 111 participants responded to at least one question.

The wearable dataset also provides a wealth of information, including: pulse rate statistics, body temperature, stand hours, walking statistics, calories burned, noise exposure. Of these, 79 participants contributed data to at least one category. Additionally, sleep data was collected in a separate dataset (presumably using a specialized sensing device) from 10 participants. PPG information, also collected in a separate dataset, was recorded from 185 participants.

### 3.2 Method Overview

Our methodology is broken up into three categories: data preprocessing, feature extraction, and classification. The full workflow can be seen in Figure 1 and a summary of the participants at each stage of processing is shown in Table 2.

Table 2. Dataset breakdown before and after processing.

| | | | | |
|---|---|---|---|---|
| **Breakdown** | | | | |
| *Source* | Automatic Sources | | | Surveys |
| *Categories* | Heart Rate Variability | Sleep | Wearables | Health |
| *Participants* | N=185 (100%) | N=10 (5%) | N=79 (42%) | N=111 (60%) |
| **Processing** | | | | |
| *Action* | Daily averaging and min-response filtering | Isolate sleep duration | Isolate pulse, step, and calorie data | Filter for COVID-19 questions, map, and envelope |
| *Participants* | N=60 (32%) | N=7 (4%) | N=32 (17%) | N=100 (54%) |
| **Training/Testing Combinations** | | | | |
| *Split* | COVID-19 + HRV | COVID-19 + Wearables | | COVID-19 + HRV + Wearables |
| *Participants* | N=45 (24%) | N=26 (14%) | | N=21 (11%) |

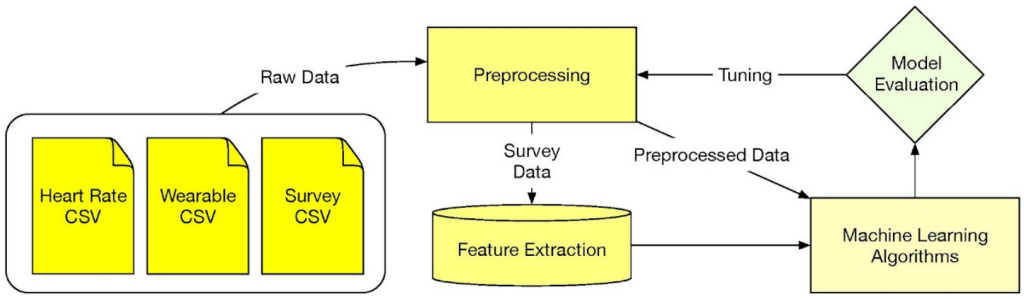


Fig. 1. Flowchart of Methodology and Approach for Analysis From Raw Data to Output Process.

*3.2.1 Data Preprocessing.* As shown in Figure 1, the Welltory dataset [9] is split into several categories including: heart rate, sleep, and wearable datasets. Table 2 further shows the breakdown of each dataset after processing.

Across all datasets, we require participants to have at least 4 data points for a given feature. This requirement, as shown in the following section, is a result of our recurrent neural network's (RNN) 4 day sliding window.In our RNN, we opted for a sliding window where tuples are grouped within a window that slides across the time interval of when a patient contracts COVID-19. Maintaining this requirement for both the RNN and traditional machine-learning algorithms allows us to compare our results across models.

For heart rate variability (HRV), we used a naive approach to preprocess it. We average the heart rate entries for each patient and drop participants with empty data or ones that do not fulfill our requirement of having at least 4 data point entries. After processing the HRV data, we are left with 60 participants.

Investigating the wearables dataset, we identified features that may be indicative of COVID-19 onset, including step count, step speed, basal calories burned, total calories burned, and wearable heart pulse information. For pre-processing this dataset, we dropped excluded sparse features including resting pulse and average body temperature. In total, there were 79 participants in the processed wearables dataset.

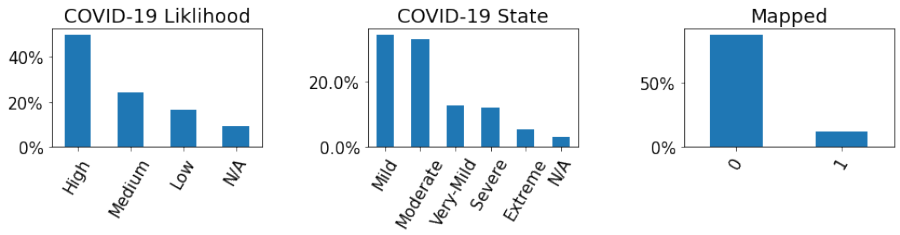


Fig. 2. Distribution of COVID-19 survey responses and resulting mapped value distribution.

In terms of sleep data, we specifically investigated sleep duration which we believe could provide insight into COVID-19 onset. Unfortunately, the sleep data was only provided by 10 participants which was further reduced to 7 after preprocessing. Given this small sample size, we decided to drop the sleep analysis entirely and instead focus on the more robust categories.

*3.2.2 Feature Extraction.* While the majority of features are simply preprocessed, the COVID-19 survey data needs to be further reduced to a workable form, i.e., a numerical representation of the likelihood of COVID-19 status. To accomplish this, two survey questions are isolated from the 11 survey questions, specifically "Overall state" and "An assessment of the user's symptoms which shows how likely it is that the user has coronavirus." The first question is scaled from 1 to 6, 1–3 indicating mild or no symptoms, and 4–6 indicating moderate to extremely sever symptoms. The latter question is scored from 1 to 5, 1 and 2 defined as having symptoms associated with COVID-19, and 3–6 defined as not having COVID-19-related symptoms. Instances where the "overall state" is between 1–3 are mapped to 0, and 4–6 are mapped to 1. Similarly, instances where the "user's symptoms which shows how likely it is that the user has coronavirus" are 3–6 are mapped to 0, and 1–2 mapped to 1. The distribution of these responses can be seen in Figure 2.

Once the COVID-19 data is mapped to a binary value, all non-zero responses are enveloped to provide a continuous indication of COVID-19 status. In other words, all days in between the first reported positive COVID-19 status and the last reported COVID-19 status are considered positive statuses, allowing us to indicate the onset and offset time for the illness. This methodology can be further refined by re-mapping the binary values to floating point numbers (e.g., medium-likelihood statuses are mapped to 0.5) and computing the sum or day-by-day multiplication rather than a binary *AND* operation.

*3.2.3 Classification.* Each dataset was fed into the machine learning models separately. The HRV data and wearables data were individually processed, and then our machine learning models were trained on those datasets. After analyzing the results, we combined the datasets and reran our analysis. The idea was to have a wider variety of features to find correlations between the inputs and the outputs.

With the binary mapping of the COVID-19 data, we are able to take our processed features and input them into a variety of classification methods. We explore three types of algorithms for this objective, namely linear regression, multi-layer perceptrons, and supper vector machines. Prior work has shown success with support vector machines (SVMs) in particular [9, 11] which we validate in our study.

Our first algorithm is Logistic Regression, a linear classifier that uses a sigmoid function to fit the training features to the output classes. The next algorithm is a Multi-Layer Perceptron (MLP), which is essentially a small neural network. Instead of deep hidden layers of neurons, the MLP simply layer several perceptrons to achieve a nonlinear classification system. This is the major difference between Logistic Regression and the MLP, and a key component that we want to explore in these experiments.

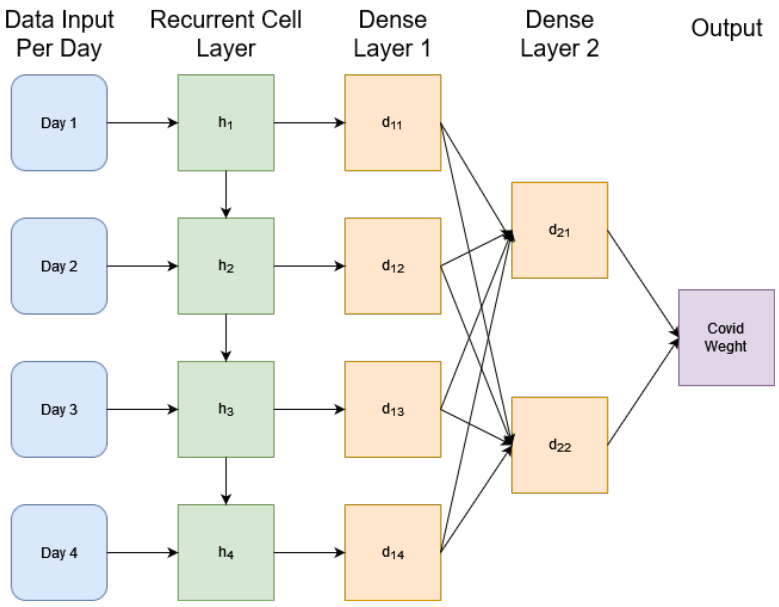


Fig. 3. RNN Setup for Predicting COVID Onset

Our remaining classification algorithms all fall under the broad category of Support Vector Machines (SVMs). Another series of classification methods, SVMs function by attempting to "separate" the input features in an $n$-dimensional space, where $n$ is the number of types of features in the data. With linear kernels, the default for SVMs, the model will then work towards creating a line to separate the two classes in our input. SVMs can also use nonlinear kernels, giving us another comparison similar to Logistic Regression and the MLP. Using nonlinear classification, it is possible to find correlations that exist in a higher dimensionality. These correlations can be found without requiring any altering of the input data, by instead changing the model to process the input data into a hyperplane and separating the two classes in that higher dimensional space.

*3.2.4 Deep Learning Model.* Classical machine learning methods have historically under-performed deep learning methods for a variety of reasons [10]. Among them is a fundamental property of deep learning models, the automatic learning of weights for feature hierarchies. With classical machine learning, the responsibility for filtering data and assigning weights to different feature columns is on the data scientist. Deep learning models, through the process of back propagation, are able to learn these weights on their own.

Since one of our goals was to predict the onset of COVID over time, we decided to build our deep learning model as a Recurrent Neural Network (RNN). Our RNN consisted of four days of feature inputs, each of which was structured as a 2D matrix. Those inputs were fed into a RNN cell layer, where the weights would begin training to determine correlation between those input values and our output (Fig 3). Then there are several dense layers, where the outputs from the RNN cell layer are trained as a full set of inputs, and then our output is predicted at the end.

The primary difference between RNNs and other deep learning models is the how the weights are trained in the first hidden layer. In a simple neural network, all the layers look like our dense layers, and the data inputs are evaluated without regard to sequencing. The RNN cell layer, with weights propagating sideways in addition to

| Algorithm | Precision | Recall | F1 | Accuracy | AUC |
|---|---|---|---|---|---|
| Logistic Regression | 0.620 | 0.640 | 0.630 | 0.631 | 0.631 |
| MultiLayer Perceptron | 0.556 | 0.683 | 0.613 | 0.576 | 0.578 |
| SVM-LinearSVC | 0.522 | 0.658 | 0.582 | 0.537 | 0.539 |
| SVM-NuSVC | 0.532 | 0.571 | 0.551 | 0.543 | 0.543 |
| SVM-SVC (linear kernel) | 0.613 | 0.658 | 0.635 | 0.628 | 0.629 |
| SVM-SVC (rbf kernel) | 0.617 | 0.621 | 0.619 | 0.625 | 0.625 |

Table 3. Prediction using Heart Rate Variability data.

| Algorithm | Precision | Recall | F1 | Accuracy | AUC |
|---|---|---|---|---|---|
| Logistic Regression | 0.558 | 0.309 | 0.397 | 0.569 | 0.550 |
| MultiLayer Perceptron | 0.409 | 0.553 | 0.471 | 0.426 | 0.436 |
| SVM-LinearSVC | 1.000 | 0.053 | 0.101 | 0.564 | 0.527 |
| SVM-NuSVC | 0.537 | 0.691 | 0.605 | 0.583 | 0.591 |
| SVM-SVC (linear kernel) | 0.493 | 0.393 | 0.438 | 0.534 | 0.524 |
| SVM-SVC (rbf kernel) | 0.349 | 0.160 | 0.219 | 0.475 | 0.453 |

Table 4. Prediction using Wearable data.

forwards/backwards, allows us to build weights around the order of our input sequences. The idea with this is to automatically detect trends between multiple days of measurements and the COVID diagnosis over those days.

Our output for the RNN was changed slightly as well, but only superficially. The filtering for each COVID datapoint remained the same, but we changed from a binary classification problem to regression problem, while still retaining the actual values of the data. Since we wanted one datapoint as an output, we could no longer use a 0 or 1 to represent four days of outputs. Ergo, we changed to a weighted system, where we average the COVID scores over the four days for a single datapoint. For example, if a subject had three days of no COVID and one day of COVID, their weighted score would be 0.25. With this altered weighting system, our RNN has the potential to look at multiple days of input data and predict how many of those days the subject had COVID.

## 4 RESULTS

The main goal of our study is to evaluate the accuracy of machine learning models in predicting COVID-19 onset. To answer this question, we construct 6 traditional machine learning models (Section 3.2.3) and a deep RNN model (Section 3.2.4). We evaluate the performance under three scenarios: utilizing HRV data; utilizing wearables data; utilizing the combined HRV and wearables data.

### 4.1 Traditional Machine Learning Models

5-fold cross-validation is developed to evaluate the traditional machine learning models. Table 3 shows the performance of the 6 models using heart rate variability data as input, Table 4 shows the result using wearable sensor data, and Table 5 shows the result using both heart rate variability data and wearable data.

Support vector machine with a linear kernel performs the best for heart rate variability data with an F1 score of 0.635 (Fig. 4). This shows that there is a high correlation between heart rate variability and COVID-19 onset which aligns with previous studies [5, 7, 11, 14]. On the other hand, the model trained with heart rate variability data outperforms the other model with wearable data. The best performing model is SVM-NuSVC with an F1 score of

| Algorithm | Precision | Recall | F1 | Accuracy | AUC |
|---|---|---|---|---|---|
| Logistic Regression | 0.677 | 0.724 | 0.7 | 0.650 | 0.640 |
| MultiLayer Perceptron | 0.627 | 0.724 | 0.672 | 0.602 | 0.584 |
| SVM-LinearSVC | 0.6 | 0.569 | 0.584 | 0.543 | 0.540 |
| SVM-NuSVC | 0.646 | 0.724 | 0.683 | 0.621 | 0.607 |
| SVM-SVC (linear kernel) | 0.732 | 0.707 | 0.719 | 0.689 | 0.687 |
| SVM-SVC (rbf kernel) | 0.536 | 0.897 | 0.671 | 0.505 | 0.448 |

Table 5. Prediction using combined data.

| Data | Precision | Recall | F1 | Accuracy | AUC |
|---|---|---|---|---|---|
| Heart Rate Variability | 0.582 | 0.550 | 0.564 | 0.735 | 0.850 |
| Wearable | 0.839 | 0.578 | 0.684 | 0.650 | 0.651 |
| Combined Dataset | 0.502 | 0.667 | 0.548 | 0.889 | 0.848 |

Table 6. Prediction using our proposed deep recurrent model.

0.605 (Fig. 5). This shows that heart rate variability data may be a better indicator for predicting COVID-19 onset than with wearables. While the models using wearable data are not as performant as HRV models, an accuracy greater than 0.5 demonstrates that the wearable data may be useful in predicting COVID-19. It is interesting to point out that SVM-LinearSVC fails to learn a meaningful model using wearable data. It mostly predicts negative except for a few exceptions, which allows it to achieve 1.0 for precision and 0.053 for recall. We believe this is due to the weak relations between the wearable data and the COVID-19 onset, which makes it challenging for the machine learning models to learn.

Finally, we combine the heart rate variability data and wearable data and analyze them all together. We only extract entries when both heart rate variability data and wearable data are available for a participant on a particular day. We see a significant improvement from the earlier results and achieve an F1 score of 0.719 using SVM-SVC with a linear kernel (Fig. 6). This strongly suggests that wearable data could be a good indicator for predicting COVID-19 onset when used in conjunction with HRV data.

To visualize the performance of our model, we additionally plot the receiver operating characteristic (ROC) curve and calculated the area under the ROC curve (AUC) which measures the performance of our classifiers across all possible classification thresholds. (see Fig. 4, 5, and 6). We achieve an AUC of 0.687 for the best performing machine learning model (SVM-SVC, linear kernel) using both heart rate variability data and wearable data.

## 4.2 Deep Recurrent Model

While the traditional machine learning algorithms look at the data one date at a time, we harness the power of deep learning and analyze multiple days simultaneously to see if we can further improve our performance. Using the network mentioned in Section 3.2.4, we are able to achieve a significantly better performance than the traditional machine learning models (see Table 6).

Using solely the heart rate variability data, the deep recurrent model gives us an AUC of 0.850, which is a 35% improvement compared to 0.629 using a support vector machine. On the other hand, the F1 score is much lower at 0.564. This is because we are performing a multi-class classification here whereas the traditional machine learning models are trained for binary classification. The wearable data gives us a better result compared to the

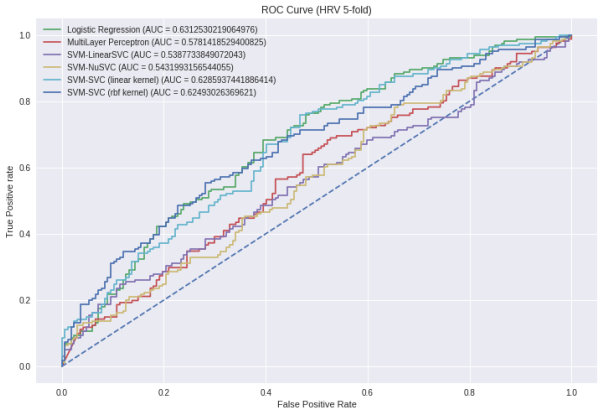


Fig. 4. ROC curve and AUC using Heart Rate Variability data

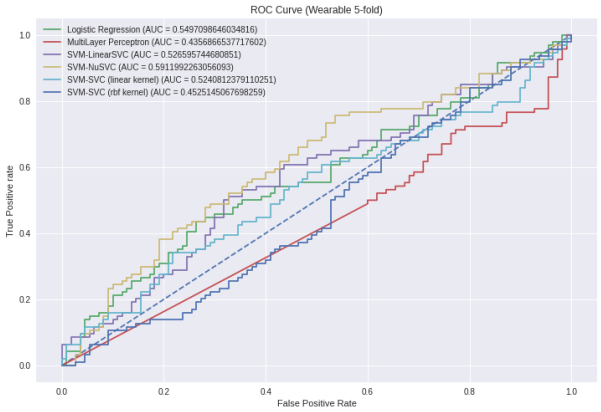


Fig. 5. ROC curve and AUC using Wearable data

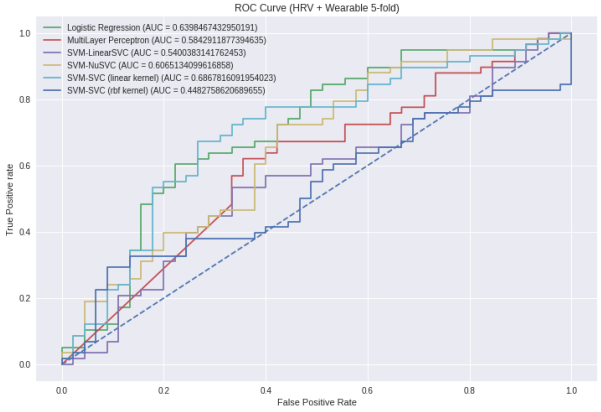


Fig. 6. ROC curve and AUC using combined data

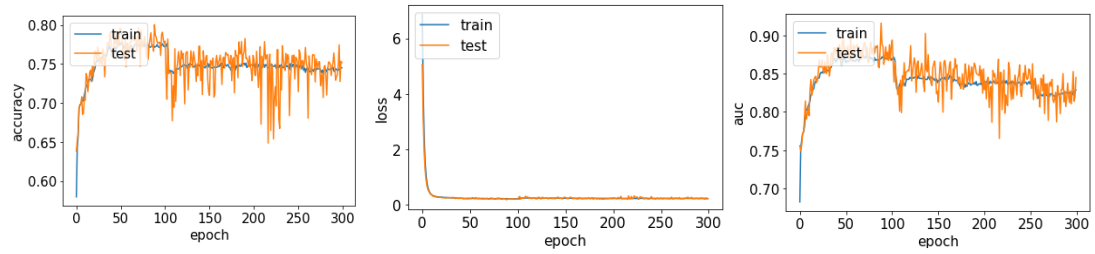


Fig. 7. RNN results with the HRV dataset

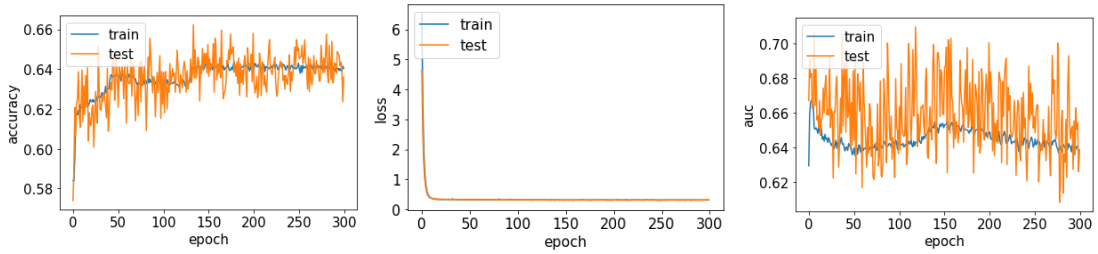


Fig. 8. RNN results with the wearables dataset

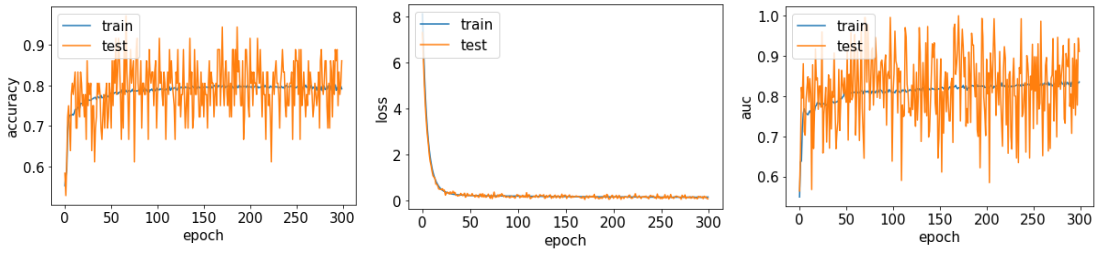


Fig. 9. RNN results with the combined dataset

traditional machine learning methods with an AUC of 0.651. This shows that the temporal information from wearable data is useful for the prediction of COVID-19 onset. When both datasets are used to train the deep model, we observe an AUC of 0.848 which is comparable to the AUC of 0.850 using the heart rate variability data alone. We believe the small discrepancy may caused by the small data size since participants do not usually have both heart rate variability data and wearable data at the same time, which gives us only 64 data points at the end after filtering. This inevitably affects the performance of the deep model since it requires a large amount of data to train.

Additionally, we plot the accuracy curve, loss curve, and the AUC curve (see Fig. 7, 8, and 9). In the accuracy plots, we can see a gradual increase in accuracy across the epochs. Interestingly, for the heart rate variability data in Fig. 7, we see a drop in performance at roughly epoch 100. We believe this is caused by the large learning rate at the later stage of the training session and can be further improved if we implement learning rate decay or fine-tune the parameters. From the loss curves, we can see our model converges quickly since we only have a simple model with a small number of features and parameters. Finally, in the AUC curves, the plots for heart rate

variability and wearable show similar trends in the heart rate variability accuracy curve. We also believe this can be improved by fine-tuning the learning rate and implementing a decayed learning rate.

## 5 DISCUSSION

In our study, we first processed the Wellroy dataset and fed it into the classical classification methods as a baseline. Our results show that using certain variations of SVM provide good F1 results of predicting COVID-19 onset for both heart rate and wearable datasets. We decided to take the study a step further by implementing an RNN model to account for temporal factors for COVID-19 onset. The results that we presented show promise in that they improved upon the classical methods. As can be seen in Table 5, the classical methods' AUC levels ranged from 0.45 to 0.69. This suggests that traditional machine learning models are decent classifiers for COVID-19 onset. When accounting for the RNN results, the AUC levels is around 0.64. While this is technically lower than the performance of the SVM-SVC (linear-kernel), we argue the model is more realistic because it predicts onset over a temporal layer – an integral aspect of a virus that doesn't merely last for a day. While not a strong indicator of classification, governments and firms could leverage using wearable data from patients to better track virus outbreaks.

Compared to existing literature, prior studies had higher AUC scores than our models but with several caveats. For example, the authors of [7] collected highly-accurate data from a smart ring wearable device called the Oura Ring.[1] Their machine learning model performed well with an AUC of 0.81, although they did not utilize temporal features and greatly smoothed the data to boost their model's performance through techniques like hot encoding. Similarly, another group [5] utilized the same dataset as this study but only investigated HRV data with a support vector machine, K-Nearest neighbor, and logistic regression classifiers. Their SVM model had a staggering AUC score of 0.93 by utilizing natural language processing techniques on the survey data to boost overall accuracy.

Despite the potential of our hypothesis, there are numerous limitations in our study. *First*, the publicly available COVID-19 dataset [9] is limited in terms of participants and response rates for essential questions. Since [9] is one of the few publicly available datasets that includes both COVID-19 data and sensing data, future work should prioritize expanding data collection. Additionally, it is imperative that these datasets be made public – either for additional study or validation – unlike other datasets that are locked under company NDAs (e.g., [12]). *Second*, alternative methods of preprocessing the COVID-19 survey data should be considered. Although we offer an intuitive mapping between the self-reported COVID-19 data, existing literature should be more thoroughly analyzed to provide a rigorous mapping between self-reported data and statistically significant mapped valued. Additionally, including healthcare providers and clinicians in future studies would benefit this area of research, as clinically validated COVID-19 results would lead to overall better model accuracy. *Third*, a more rigorous ablation study should be conducted to identify the most useful features contributing to COVID-19 onset detection. Although this work identifies multiple features to consider that are intuitively linked to health-related sensing (e.g., changes in pulse, decrease in physical activity), a full ablation study could identify features that are statistically more significant. Additionally, this can guide future studies in improving data collection methods, allowing them to focus on specific wearable sensors. *Fourth*, but not least, this study naturally lends itself to COVID-19 offset prediction in addition to onset prediction. With finer-grained COVID-19 survey data, wearable devices could conceivably enable individuals to predict when their COVID-19 illness will end. Similar to onset detection, investigating the recovery of physiological signals could allow individuals to test out of isolation early, or at least offer them hope that the illness has almost passed.

[1]The Oura Ring collects dermal temperature, heart rate variability, respiratory rate, and estimates physical activity based on accelerometer data.

## 6 CONCLUSION

In this paper, we present our initial results at predicting the onset of COVID-19 using data from wearable devices. Utilizing a publicly available database, we validate the claim of prior work that COVID-19 onset can be predicted, with reasonable accuracy, using heart rate variability data. We expand on this claim by introducing data from additional sensors, including step count, pulse rate, and caloric data, and analyzing the performance of various machine learning models. Finally, we elaborate on future directions of the study and offer suggestions for model- and methodology-related improvements.

## REFERENCES

[1] Surekha Borra. 2020. COVID-19 apps: Privacy and security concerns. In *Intelligent Systems and Methods to Combat Covid-19*. Springer, 11–17.

[2] Paolo Boscolo-Rizzo, Daniele Borsetto, Claire Hopkins, et al. 2020. Challenges in interpreting the diagnostic performance of symptoms to predict COVID-19 status: the case of anosmia. In *International forum of allergy & rhinology*. Wiley-Blackwell.

[3] Talha Burki. 2022. Dynamic zero COVID policy in the fight against COVID. *The Lancet Respiratory Medicine* (2022).

[4] A.K. Yetisen H.C. Ates and F. Güder et al. 2021. Wearable devices for the detection of COVID-19. *Nat Electron* 4 (2021), 13–14. https://doi.org/10.1038/s41928-020-00533-1

[5] Haytham Hijazi, Manar Abu Talib, Ahmad Hasasneh, Ali Bou Nassif, Nafisa Ahmed, and Qassim Nasir. 2021. Wearable Devices, Smartphones, and Interpretable Artificial Intelligence in Combating COVID-19. *Sensors* 21, 24 (2021), 8424.

[6] Haitham Jahrami, Ahmed S BaHammam, Nicola Luigi Bragazzi, Zahra Saif, MoezAlIslam Faris, and Michael V Vitiello. 2021. Sleep problems during the COVID-19 pandemic by population: a systematic review and meta-analysis. *Journal of Clinical Sleep Medicine* 17, 2 (2021), 299–313.

[7] Ashley E Mason, Frederick M Hecht, Shakti K Davis, Joseph L Natale, Wendy Hartogensis, Natalie Damaso, Kajal T Claypool, Stephan Dilchert, Subhasis Dasgupta, Shweta Purawat, et al. 2022. Detection of COVID-19 using multimodal data from a wearable device: results from the first TemPredict Study. *Scientific reports* 12, 1 (2022), 1–15.

[8] Su HW. Natarajan A. and C. Heneghan. 2020. Assessment of physiological signs associated with COVID-19 measured using wearable devices. *npj Digit. Med.* 3 (2020), 156. https://doi.org/10.1038/s41746-020-00363-7

[9] Pavel. 2021. COVID-19 and Wearables Open Data Research. osf.io/zd7gq

[10] A Picon, A Alvarez-Gila1, U Irusta, and J Echazarra. 2020. Why Deep Learning Performs Better than Classical Machine Learning. *Ingenieria e Industria* (2020), 119–122. https://doi.org/10.6036/9574

[11] Alexey Ponomarev, Konstantin Tyapochkin, Ekaterina Surkova, Evgeniya Smorodnikova, and Pavel Pravdin. 2021. Heart rate variability as a prospective predictor of early COVID-19 symptoms. *medRxiv* (2021).

[12] M Poongodi, Mounir Hamdi, Mohit Malviya, Ashutosh Sharma, Gaurav Dhiman, and S Vimal. 2021. Diagnosis and combating COVID-19 using wearable Oura smart ring with deep learning methods. *Personal and ubiquitous computing* (2021), 1–11.

[13] M. Wang T. Mishra and A.A. Metwally et al. 2020. Pre-symptomatic detection of COVID-19 from smartwatch data. *Nat Biomed Eng* 4 (2020), 1208–1220. https://doi.org/10.1038/s41551-020-00640-6

[14] Gatha Tanwar, Ritu Chauhan, Madhusudan Singh, and Dhananjay Singh. 2020. Pre-emption of affliction severity using HRV measurements from a smart wearable; case-study on SARS-CoV-2 symptoms. *Sensors* 20, 24 (2020), 7068.

[15] Alma Tostmann, John Bradley, Teun Bousema, Wing-Kee Yiek, Minke Holwerda, Chantal Bleeker-Rovers, Jaap Ten Oever, Corianne Meijer, Janette Rahamat-Langendoen, Joost Hopman, et al. 2020. Strong associations and moderate predictive value of early symptoms for SARS-CoV-2 test positivity among healthcare workers, the Netherlands, March 2020. *Eurosurveillance* 25, 16 (2020), 2000508.

[16] Souheil Zayet, Timothee Klopfenstein, Julien Mercier, Ludovic Lan Cheong Wah, Pierre-Yves Royer, Lynda Toko, Vincent Gendrin, et al. 2021. Contribution of anosmia and dysgeusia for diagnostic of COVID-19 in outpatients. *Infection* 49, 2 (2021), 361–365.