An Adaptive Framework for COVID-19 Prognosis through Online Deep Learning

Haotian Chen
University of Illinois Urbana-Champaign, USA

Yogatheesan Varatharajah
University of Illinois Urbana-Champaign, USA

Yurui Cao
University of Illinois Urbana-Champaign, USA

Krishnakant Saboo
University of Illinois Urbana-Champaign, USA

Karl M. Kochendorfer
College of Medicine, University of Illinois at Chicago, USA

Andrew B. Trotter
College of Medicine, University of Illinois at Chicago, USA

William L. Galanter
College of Medicine, University of Illinois at Chicago, USA

Ravishankar K. Iyer
University of Illinois Urbana-Champaign, USA

ABSTRACT

With disease outcomes constantly being impacted by virus variants, enhanced treatments, and widespread vaccination, the fast-changing COVID-19 pandemic demands adaptability in prognostic systems. This paper presents an online deep-learning-based framework for predicting severe disease outcomes (including intensive care unit [ICU] admission and death) among hospitalized COVID-19 patients. The proposed framework starts with an initial prognostic model based on an artificial neural network and continues to update it using data from new patients. We applied our framework on a dataset of 595 patients collected over 16 months. We developed an initial model using 4 months of data, and then retrospectively evaluated it in an online fashion using the data over a 1-year period. Our method achieved an AUC of 0.762 in predicting severe outcomes, 7% better than existing COVID-19 prognostic methods and 6% better than benchmark online-learning methods. The obtained prediction accuracy is also superior to that of model retraining.

CCS CONCEPTS

• **Applied computing** → **Health care information systems**; • **Computing methodologies** → *Online learning settings*;

KEYWORDS

COVID-19, Disease Prognosis, Adaptive, Online Machine Learning

1 INTRODUCTION

The COVID-19 pandemic has posed unprecedented challenges to modern healthcare systems. Given the substantial variability in severity of infection, it is important for hospitals to predict whether a patient will develop severe outcomes. Early identification of high-risk patients facilitates early interventions and optimal allocation of limited medical resources [15]. Various prognostic models have been developed for predicting severe disease progression. During the initial phase of the pandemic, researchers developed generalized linear models, such as logistic regression (LR). Subsequently, ensemble models such as random forest (RF) [14] and extreme gradient boosting (XGBoost) [8] were utilized to characterize nonlinear patterns in disease progression; the ensemble strategy makes these models less susceptible to over-fitting. Recently, deep learning models have been proposed to learn more complex patterns using large patient cohorts [10, 15].

The existing literature mainly focused on demonstrating the feasibility of various machine learning methods in predicting COVID-19 progression. However, very little effort has been dedicated to developing adaptive machine learning frameworks that can continuously update model parameters as the pandemic evolves. This is particularly important because disease outcomes are continuously impacted by virus variants, widespread vaccination, and improved treatments [3, 4]. It is widely understood that new data must be incorporated into models in a timely fashion to enable reliable prognoses throughout the pandemic.

Online machine learning (ML) allows continuous re-calibration of ML models in fast-changing environments and has been useful in domains such as computer security and power systems [9, 12]. In light of the continuous streams of patients admitted to hospitals and the need to adapt to the evolving disease, an online ML paradigm appears to be especially appropriate for the fast-changing COVID-19 pandemic. Hence, we developed and evaluated an online ML framework for COVID-19 prognosis using an artificial neural network (ANN) as the base model. On top of the conventional online ML paradigm, we introduced a primary-backup mechanism to minimize modifications of the model in use. The *primary* model performs real-time prognostications while the *backup* is updated continuously, and the primary is replaced with the backup only when a performance degradation is detected.

We conducted a retrospective analysis using 16 consecutive months of clinical data of 595 patients who were hospitalized at the University of Illinois Health System (UIH) in Chicago, Illinois. We trained an initial model using a *development* cohort (3/1/20–6/30/20) and then evaluated the model in an online fashion using an *evaluation* cohort (7/1/20–6/30/21). Our results demonstrate superior robustness over both existing COVID-19 prognostic methods and benchmark online learning methods. In an upcoming yet highly uncertain endemic phase, the proposed approach will be continuously essential for capturing changing circumstances. More broadly, as a lesson learned from the COVID-19 patient care, this work advocates an integration of online machine learning into traditional disease outcome modeling for an up-to-date prognosis.

2 METHODOLOGY

Here we describe the online deep learning-based framework that includes a primary ANN for outcome prediction and a backup ANN for model update (Figure 1).

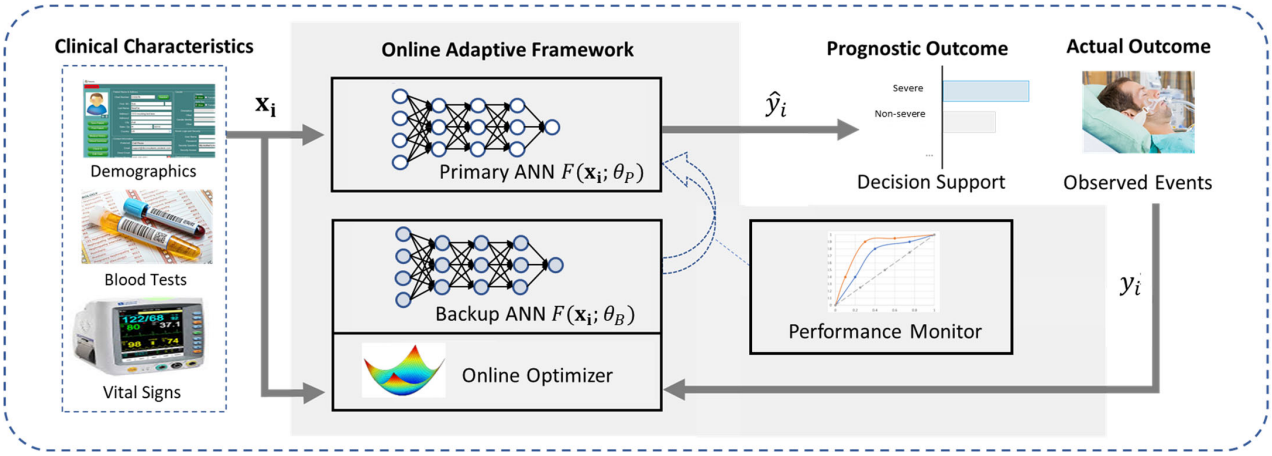

Figure 1: Online deep-learning-based framework to provide up-to-date prognosis for individual patients.

2.1 Overall analytic flow

As illustrated in Figure 1, when a new patient is admitted to the hospital, his/her clinical characteristics are collected. We use $\mathbf{x_i}$ to denote a vector containing the attributes of patient i, which include age, albumin level, white blood cell count, and SpO2 value, as detailed in Section 3.1. The primary ANN takes $\mathbf{x_i}$ as input and produces an outcome $\hat{y}_i$, a binary variable indicating whether the patient will enter a severe stage (ICU admission or death) at any point during hospitalization.

When the ground truth outcome is subsequently observed, it is used to update the model parameters of the backup ANN. We use y_i to denote the observed clinical outcome of patient i. The predicted and ground truth outcomes are tracked by a *performance monitor*. When performance degradation is detected, the primary ANN is replaced with the backup.

2.2 Framework components

Our framework consists of four components: a primary ANN, a backup ANN, an online optimizer, and a performance monitor.

The **primary ANN** $F(\mathbf{x_i}; \theta_P)$ takes patient measurements as inputs to predict patient outcomes. The ANN is configured with three fully connected hidden layers. Each layer has three units with sigmoid activation functions. Given a new patient with characteristics $\mathbf{x_i}$, the primary ANN predicts his/her probability of severe outcome as $P(y_i = 1|\mathbf{x_i}) = F(\mathbf{x_i}; \theta_P)$. We use θ_P to denote the weights and biases of the primary ANN. θ_P is trained initially using the development cohort.

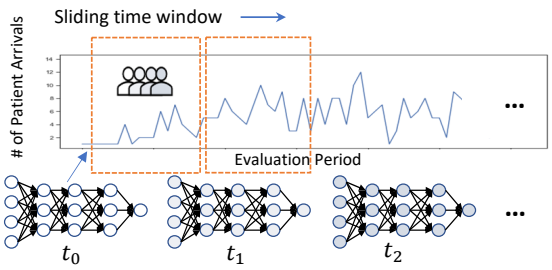

Figure 2: Progressive update of the backup ANN.

The **backup ANN** $F(\mathbf{x_i}; \theta_B)$ is an ANN being updated in the background by an online optimizer. θ_B is initialized with the parameters trained using the development data, same as the initial θ_P. Later at the evaluation cohort, θ_B is continuously updated using new patient data within the latest time window, as the time window periodically slides forward (illustrated in Figure 2).

The **online optimizer** updates the backup model parameters θ_B after a time window W has passed. This optimizer, denoted as GD, uses gradient descent and minimizes a cross-entropy loss function:

$$l(\mathbf{x_i}, y_i; \theta_B) = -y_i \log F(\mathbf{x_i}; \theta_B) - (1 - y_i)\log(1 - F(\mathbf{x_i}; \theta_B)) \quad (1)$$

We formed a mini-batch using the newly collected data $\{(\mathbf{x_i}, y_i)\}$ and performed mini-batch gradient descent updates N times to minimize the effect of noise from individual samples. Here W and N are hyperparameters. Denoting the size of $\{(\mathbf{x_i}, y_i)\}$ as M:

$$\theta_B \leftarrow \theta_B - \frac{\eta}{M} \sum_{i=1}^{M} \frac{\delta l(\mathbf{x_i}, y_i; \theta_B)}{\delta \theta_B} \quad (2)$$

The **performance monitor** detects the degradation of the in-use ANN and triggers a replacement. The monitor compares the areas under receiver operating characteristic curves (AUCs) within the current time window with the previous one. (If prediction accuracy decreases, the *Monitor* function returns *True* in Algorithm 1.)

The execution in each time window is summarized in Algorithm 1. The forward pass for outcome prediction and the backward pass for model update can be implemented in two independent threads, because of the isolation between the in-use ANN and backup ANN.

3 EXPERIMENTS

Using the data collected at UIH, we evaluated a one-year-long performance of the proposed approach.

3.1 Patient data

The data included demographics, blood tests, and vitals of a total of 595 COVID-19 patients admitted to UIH between 3/1/20 and 6/30/21. According to [2], we selected the predictive and easily collectible features, including age, albumin level, white blood cell count, and

Algorithm 1 Online update

Input: θ_P, θ_B, streaming patient data $\{(\mathbf{x_i}, y_i)\}$ within the time window W
Output: θ_P, θ_B
Initialization:
$S_p \leftarrow \emptyset$ /* S_p for storing predictive outcomes */
$S_g \leftarrow \emptyset$ /* S_g for storing ground truths */
for i in patient flow **do**
  $\hat{y_i} \leftarrow F(\mathbf{x_i}; \theta_P)$
  S_p.insert($(\mathbf{x_i}, \hat{y_i})$)
  S_g.insert($(\mathbf{x_i}, y_i)$)
end for
if $Monitor(S_p, S_g)$ is True **then**
  $\theta_P \leftarrow \theta_B$ /* Replace the primary ANN */
end if
$\theta_B \leftarrow GD(\theta_B, S_g)$

SpO2 value, as inputs to the model. Missing data was imputed with the population mean of the development cohort.

3.2 Model training and evaluation

To represent model development and deployment during the pandemic, we divided the data into two cohorts by patient admission date: 1) **development cohort** (3/1/20–6/30/20) and 2) **evaluation cohort** (7/1/20–6/30/21), where the one-year evaluation cohort includes the emergence of the Delta variant and mRNA vaccines. Model performance during each quarter of the evaluation period is used to show temporal variations in prediction performance. Traditional offline models were trained using the development cohort, while online models were trained using the development cohort and then updated in the evaluation cohort.

We evaluated the performance using AUC, a commonly used metric for binary classification. We compared the proposed method with 1) existing ML models adopted for COVID-19 prognosis, including LR, RF, XGBoost, and offline ANN; and 2) benchmark online learning methods that are applicable but have not yet been studied for COVID-19 prognosis, including online LR and adaptive RF [5]. We chose N to be 500 and the window size W to be 3 months (i.e., quarters) based on a grid search, which are discussed in Section 5.3.

4 RESULTS

The predictive performances of the proposed approach and three popular offline COVID-19 prognostic models are shown in Figure 3, for each quarter in the evaluation period. Both LR and RF show

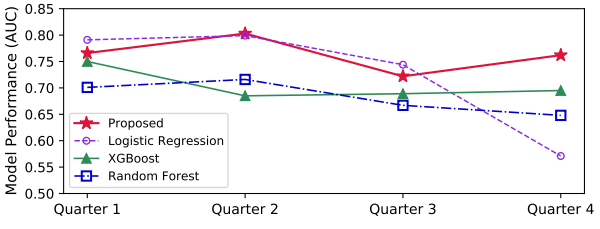

Figure 3: Comparison of the proposed model versus popular offline prognostic models in the one-year evaluation period.

Table 1: Comparison with different offline and online learning methods.

Approach	Online	AUC at evaluation phase		
		Average	First Quarter	Last Quarter
Proposed	Yes	**0.763**	0.766	**0.762**
Adaptive RF [5]	Yes	0.714	0.734	0.705
Online LR	Yes	0.736	**0.824**	0.620
ANN [6]	No	0.725	0.766	0.610
RF [14]	No	0.639	0.724	0.695
XGBoost [8]	No	0.705	0.750	0.695
LR	No	0.725	0.786	0.571

evident degradation trends during this time frame. XGBoost has a sharp accuracy drop in the second quarter and maintains the deteriorated performance thereafter. By contrast, despite fluctuations, the proposed method is able to maintain a robust AUC above 0.7 and yields the highest AUC in the last quarter.

A complete comparison with existing offline COVID-19 prognostic models and potentially applicable online learning models is presented in Table 1. For the average AUC over four quarters, the proposed model (0.763) is 4% higher than the best offline models (0.725) and 3% higher than the best online benchmark (0.736). The differences are larger in the last quarter, in which the AUC of the proposed model (0.762) is 7% higher than that of the best offline model (0.695) and 6% higher than that of the best online benchmark (0.705). An overall comparison between the offline and online learning models shows the priority of online methods.

With increasing data, a simpler model update strategy is to retrain with all available data. Figure 4 presents the results of a controlled experiment where the network structures were identical, but the parameters were optimized in offline, retraining, and online fashions respectively. Results show that the retrained ANN performs similarly to the original offline ANN. We believe that it is because the data were accumulated from the past and new data only accounted for a small portion. By contrast, the online learning approach is more responsive to recent data.

5 DISCUSSION

In this section, we discuss the changes in clinical characteristics over the pandemic, which demonstrate the necessity of online models. Furthermore, we discuss the benefits of a primary-backup mechanism and the effects of hyperparmeters.

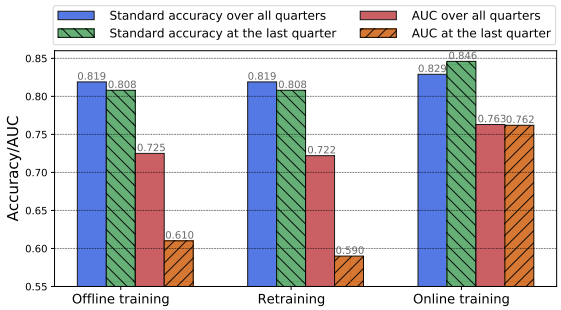

Figure 4: Performance of ANNs trained in different ways.

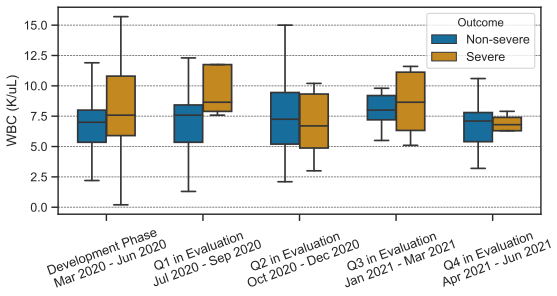

Figure 5: Relationship of WBC and disease outcome in different periods.

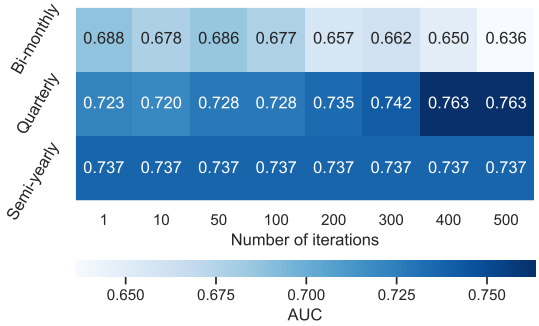

Figure 6: Prediction performance in different hyperparameter settings.

5.1 Disease changes to be captured

Figure 5 shows the distribution of WBC count over time. In the initial phase of the pandemic, worse outcomes were associated with higher WBC counts as reported in [16]. However, in the last quarter, the WBC counts of patients with severe outcomes were relatively lower, as previously reported by a study on the Delta surge [7]. Because of the diminishing similarity of later patients to the development phase, the initially trained ANN substantially degraded in the last quarter (AUC=0.610) (Table 1). While clinical reasons for this change are unclear, the varying distributions in patient characteristics warrants an adaptive framework like the one proposed in this work and others [13].

5.2 Benefits of configuring a backup ANN

The backup ANN is essentially a set of parameters being updated in the background. Compared to a direct update on the primary ANN, this design isolates the update of parameters, allowing three benefits in practice: 1) It guarantees efficiency when there is a large feature space, a deep network structure, and massive patient data, since model updates can happen in the background and will not interrupt the model in use. 2) It avoids unnecessary changes. If the current model is already performing well, it is risky to replace it with a new one that might perform worse. In this regard, our approach replaces the primary model when it exhibits performance degradation. 3) It allows regulatory reviews before model update. Many studies [1, 11] argue that ML models should be cautiously validated before clinical use, which discourages direct changes to the model in use. The proposed framework allows validation of the backup model before replacement.

5.3 Selection of hyperparameters

The predictive performance of the proposed method is influenced by two hyperparameters: 1) window size W, and 2) number of iterations N of update performed by the online optimizer. We conducted an experiment to evaluate the variability in prediction performance with different W and N values.

Figure 6 shows the results of the hyperparameter analysis. With small W (i.e., bi-monthly update), increasing N caused an overfitting to new data and a drop in AUC. However, increasing W to 3 months (quarterly update) mitigated this effect. We achieved the best performance with quarterly updates when N was greater than 400. On the other hand, we did not see any variability in performance with semi-yearly (6 months) updates.

6 CONCLUSION

This paper introduced an online deep learning-based framework for predicting severe COVID-19 outcomes at hospital admission. Using the data from a large hospital in Chicago, we demonstrated better performance than benchmark online and offline learning approaches over a one-year evaluation period. The modular structure of the proposed framework minimizes interruptions to the primary ANN. Going forward we will assess the value of our framework in other infectious disease.

REFERENCES

[1] Bibb Allen et al. 2021. Evaluation and Real-World Performance Monitoring of Artificial Intelligence Models in Clinical Practice: Try It, Buy It, Check It. *Journal of the American College of Radiology* 18, 11 (2021), 1489–1496.

[2] Haotian Chen, , et al. 2020. A Retrospective Longitudinal Study of COVID-19 as Seen by a Large Urban Hospital in Chicago. *medRxiv* (2020). https://doi.org/10.1101/2020.11.29.20240606

[3] Owen Dyer. 2021. Covid-19: Unvaccinated face 11 times risk of death from delta variant, CDC data show. *BMJ* 374, n2282 (2021).

[4] Shikha Garg et al. 2021. Clinical trends among US Adults hospitalized with COVID-19, March to December 2020: A cross-sectional study. *Annals of Internal Medicine* 174, 10 (2021), 1409–1419.

[5] Heitor M Gomes et al. 2017. Adaptive random forests for evolving data stream classification. *Machine Learning* 106, 9 (2017), 1469–1495.

[6] Thao Thi Ho et al. 2021. Deep learning models for predicting severe progression in COVID-19-infected patients: Retrospective study. *JMIR Medical Informatics* 9, 1 (2021), e24973.

[7] Kaiyuan Hu et al. 2021. COVID-19: Risk factors for severe cases of the Delta variant. *Aging (Albany NY)* 13, 20 (2021), 23459.

[8] Jonathan Montomoli et al. 2021. Machine learning using the extreme gradient boosting (XGBoost) algorithm predicts 5-day delta of SOFA score at ICU admission in COVID-19 patients. *Journal of Intensive Medicine* 1, 2 (2021), 110–116.

[9] Abdurrahman Pektaş et al. 2015. Runtime-behavior based malware classification using online machine learning. In *2015 World Congress on Internet Security (WorldCIS)*. IEEE, 166–171.

[10] H Swapna Rekha et al. 2021. Deep learning for COVID-19 prognosis: A systematic review. In *Intelligent Computing in Control and Communication*. Springer, 667–687.

[11] Wojciech Samek et al. 2019. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. LNCS, Vol. 11700. Springer.

[12] Rob Shipman et al. 2021. Online Machine Learning of Available Capacity for Vehicle-to-Grid Services during the Coronavirus Pandemic. *Energies* 14, 21 (2021).

[13] Yogatheesan Varatharajah et al. 2020. A Dynamic Human-in-the-loop Recommender System for Evidence-based Clinical Staging of COVID-19. In *HealthRecSys@ RecSys*. 21–22.

[14] Jie Wang et al. 2020. A descriptive study of random forest algorithm for predicting COVID-19 patients outcome. *PeerJ* 8 (2020), e9945.

[15] Laure Wynants et al. 2020. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 369, m1328 (2020).

[16] Bin Zhu et al. 2021. Correlation between white blood cell count at admission and mortality in COVID-19 patients: A retrospective study. *BMC Infectious Diseases* 21, 1 (2021), 1–5.