Improving ECG-based COVID-19 diagnosis and mortality predictions using pre-pandemic medical records at population-scale

Anonymous Author(s)
Affiliation
Address
email

Abstract

Pandemic outbreaks such as COVID-19 occur unexpectedly, and need immediate action due to their potential devastating consequences on global health. Point-of-care routine assessments such as electrocardiogram (ECG), can be used to develop prediction models for identifying individuals at risk. However, there is often too little clinically-annotated medical data, especially in early phases of a pandemic, to develop accurate prediction models. In such situations, historical pre-pandemic health records can be utilized to estimate a preliminary model, which can then be fine-tuned based on limited available pandemic data. This study shows this approach – pre-train deep learning models with pre-pandemic data – can work effectively, by demonstrating substantial performance improvement over three different COVID-19 related diagnostic and prognostic prediction tasks. Similar transfer learning strategies can be useful for developing timely artificial intelligence solutions in future pandemic outbreaks.

1 Introduction

Coronavirus disease (COVID-19), the infectious disease caused by the SARS-CoV-2 virus, led to a worldwide pandemic beginning at end of 2019, with 546,363,985 confirmed cases and 6,336,802 deaths as of July 2022 [1]. Initially regarded as a respiratory infection, COVID-19 is now known to affect all major systems in the body, including the cardiovascular system by causing myocardial damage, vascular inflammation, plaque instability, and myocardial infarction [2]. COVID-19 patients have shown ischemic changes, rate, rhythm abnormalities, and conduction defects in their *electrocardiograms* (ECG) [3]. In this context, machine-learned models that use ECG data could prove helpful in recognizing patients who require urgent definitive management.

Moreover, the presence of cardiac involvement in COVID-19 is known to result in poor prognosis and adverse outcomes [4]. Hence, monitoring the cardiac function is crucial for prompt identification of the appropriate actions. The ECG, which provides important information about the structure and electrical activity of the heart, is a simple point-of-care diagnostic tool [5] that can be employed to assess cardiovascular involvement in COVID-19 patients. Despite the massive impacts of the novel coronavirus, most earlier studies employing machine learning methods to predict COVID-19 outcomes using ECGs are conducted in limited data sizes [6, 7, 8]. In general, there are no prediction models that are trained and tested on population-scale ECG data to identify who is most likely to test positive for COVID-19; and among those who have COVID-19, who is most likely to suffer adverse consequences.

In this study, we use population-scale administrative health records and large ECG datasets, with two year coverage from the start of the pandemic (January 2020 to December 2021) to develop deep

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

learning (DL) models to predict diagnostic and prognostic outcomes in COVID-19. Accordingly, we develop DL models to: (1) identify patients with COVID-19; and (2) identify COVID-19 patients who are at higher risk of mortality. Further, we demonstrate that transfer learning from pre-pandemic clinical records (Feb 2007 to Dec 2019) can be utilized to improve the prediction performance of COVID-19 models.

2 Methods

In this study, we explored three COVID-19 related prediction tasks using 12-lead ECG tracing, age and sex of patients, including:

1. Diagnosis prediction with a binary classification framework (predicting if a given patient has COVID-19 or not).
2. Mortality prediction with a binary classification for 30-day mortality (i.e. if a patient with COVID-19 will die within 30 days after the acquisition of ECG).
3. Mortality prediction on a continuous time-scale using an individualized survival distribution (ISD) for each patient with COVID-19.

Anonymous province of anonymous country has a single-payer and single-provider healthcare system. The 4.4 million residents have universal access to hospital, ambulatory, laboratory, and physician services. For this study, ECG data were linked with the provincial-scale administrative health databases using unique patient health numbers. These databases included information on inpatient hospitalizations, outpatient clinic and emergency department (ED) visits, demographic information, and date of death. We used standard ECG voltage-time series, sampled at 500 Hz for 10 seconds for each of 12 leads (500 x 10 x 12 voltage measurements per ECG). We also used age and sex of the patients as additional features. An ECG record was linked to a healthcare episode if the acquisition date was within the timeframe between the admission date and discharge date of an episode. The COVID-19 diagnosis was identified using the ICD-10 codes (U071, U073) [9] in the primary or secondary diagnosis fields of a healthcare episode linked to a particular ECG. This study received ethics approval from the Health Research Ethics Board at the anonymous institution.

Our dataset consisted of 73,842 patients during the COVID-19 pandemic period (Jan 2020 - Dec 2021) who had taken 292,937 ECG tests overall. We excluded ECGs that could not be linked to healthcare episodes, patients < 18 years of age, and ones with poor signal quality. Then, we split our ECG dataset into the development dataset (random 60%: 33,197 patients with 119,761 ECGs, used for training and internal validation) and holdout dataset (remaining 40%: 22,132 patients with 80,097 ECGs), while ensuring that ECGs from the same patient were not shared between the sets. In addition, we also used a retrospective pre-pandemic dataset that included 260,774 patients with 1,992,415 ECG obtained between Feb 2007 to Dec 2019; see Figure 1).

First, we trained our models within the pandemic time frame (Jan 2020 to Dec 2021) and then we extended our training to pre-pandemic time frame (Feb 2007 to Dec 2019) to compare the performance improvement. Since the data on COVID-19 diagnosis was not available before the pandemic, we used a model for predicting over 1414 ICD-10 codes (previously developed on pre-pandemic data [10]) as a pre-training model for the COVID-19 diagnosis task. Similarly, since data on COVID-19 related deaths were not available before the pandemic, we used the all-cause mortality prediction model (again, previously developed on pre-pandemic data, unpublished) as a pre-training model for the COVID-19 mortality task. Both of these pre-trained models were then fine-tuned on COVID-19 specific predictions using the development set (60% of pandemic data) and evaluated on hold-out set (40% of pandemic data); see Figure 1.

For the classification tasks, we implemented a convolutional neural network (CNN) based on the ResNet, consisting of a convolutional layer, 4 residual blocks with 2 convolutional layers per block, followed by a dense layer (total of 6,400,433 model parameters). We used batch normalization [11], ReLU and dropout [12] after each convolutional layer. Our architecture was based on a previously published large-scale study to identify abnormalities in 12-lead ECGs [13] with some modifications to accommodate tabular data input and binary output. Each ECG instance was loaded as a 12 x 4096 numeric matrix. Sex (binary) and age features (real valued) were passed to a 10-hidden-unit layer, then concatenated with the dense layer, and finally passed to a sigmoid function to produce the output

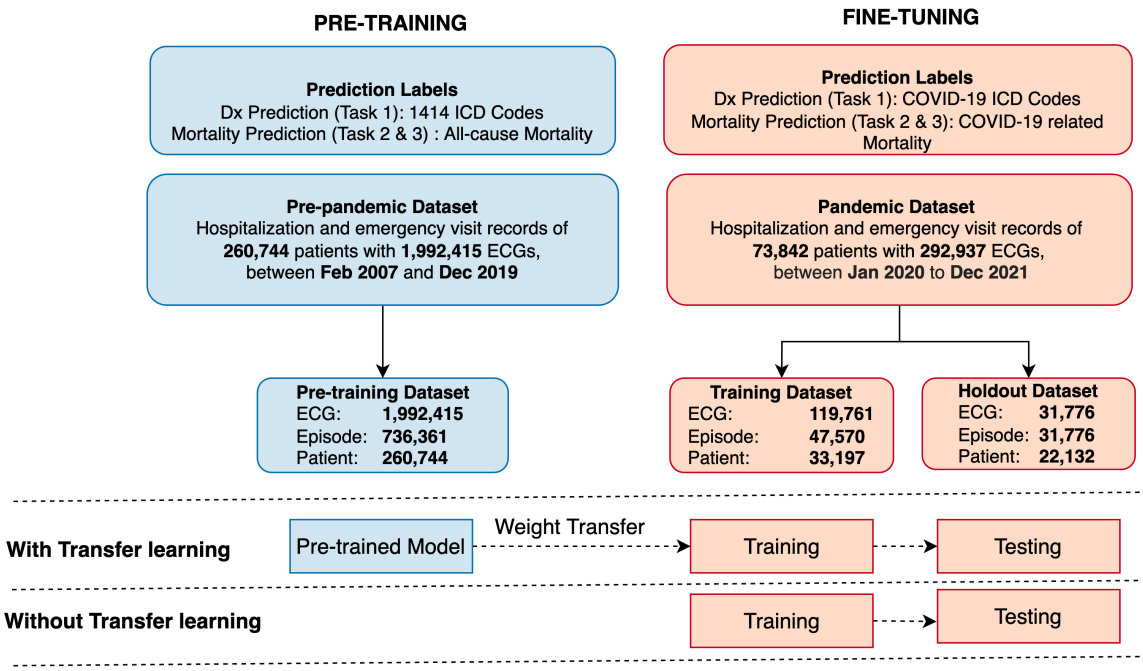

Figure 1: Datasets used for pre-training and then fine-tuning the COVID-19 prediction tasks

(Task 1: Probability of diagnosis; or Task 2: Probability of 30-days mortality). Binary cross-entropy was used as the loss function with the initial learning rate (LR) of 10^{-3}, Adam optimizer [14], ReLU activation function, kernel size of 16, batch size of 512, and dropout rate of 0.2 with other hyperparameters set to default. The LR was reduced by factor of 10 if there was no improvement in tuning loss for 7 consecutive epochs (minimum LR = $10^{-}6$), and the learning process was stopped if the loss did not reduce for nine epochs. For individual survival prediction tasks (Task 3), we designed an end-to-end ISD algorithm, which is consistent with the ResNet structure (the sigmoid was replaced with ReLu function) followed by 3 fully connected layers and a Multi Task Logistic Regression [15] (MTLR, see appendix for overview) module.

In order to implement transfer learning, (a) in the classification algorithm, we froze all previous layers' weights and added one fully connected layer and; (b) in the ISD algorithm, we froze all layers except the MTLR layers, during the training process. The models were implemented using PyTorch 1.11 in Python 3.8. We trained all our models with 8 Tesla V100-SXM2 GPUs and 32 GB of RAM per GPU.

Our goal is to ultimately generate a prediction system that could be employed at the point of care – i.e when the patient's first ECG is acquired during an ED visit or hospitalization. We therefore used the first ECG of each episode in the holdout set to evaluate our diagnosis model, and first ECG of a random COVID-19 episode for each patient in the holdout set to evaluate our mortality models. The validation performance of these models were compared using area under the receiver operating characteristic curve (AUROC) or the Concordance-index (C- index) and bootstrapped 95% confidence intervals with 1000 replicates (CI). Additionally, we report Area Under the Precision-recall Curve (AUPRC), and Average Precision (AP), F1, specificity, recall, precision, accuracy, and Brier score for binary classification as well as L1 marginal loss, and L1 hinge loss for ISD prediction.

3 Results

Supplementary Table C.1 and Figure 2 show the evaluation performance of three models, with and without pre-pandemic transfer learning. After applying the exclusion criteria, out of the total of

[1]AUROC: AUROC: area under the receiver operating characteristic curve; C-Index: concordance index; CI: Confidence Interval

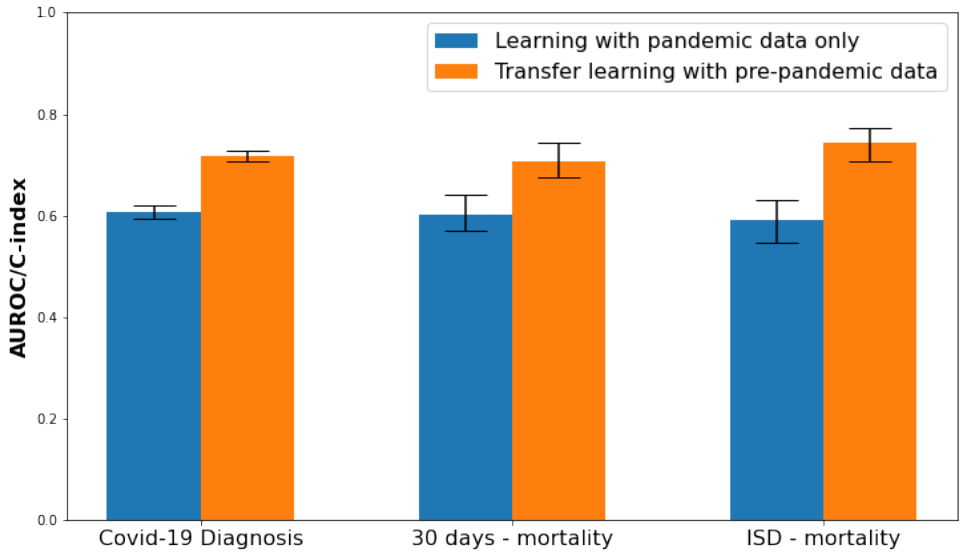

Figure 2
Comparison of AUROC / C-Index for COVID-19 prediction models with and without pre-pandemic transfer learning. Note error bars are 95%CI.[1]

199,858 ECGs acquired from 55,329 patients in the pandemic period, 3,452 (6.24%) were tested positive for COVID-19. The diagnosis classification model trained solely with a development dataset from this period showed AUROC of 60.7%. When the same algorithm was pre-trained with ICD-wide diagnosis from the pre-pandemic time frame, the COVID-19 diagnosis performance improved significantly by 11.1% to 71.8%.

Further, 407 (11.79%) patients out of the total 3,452 patients who were diagnosed with COVID-19 died within 30 days (See Kaplan Meier curve, Supplementary Figure A.3). Similar to the diagnosis task, performance of our 30-days mortality prediction improved significantly by 10.6% from AUROC of 60.1% to 70.7% when the model was pre-trained with pre-pandemic all-cause 30-day mortality.

Predicted survival curves using our ISD model for representative patients with COVID-19 are shown in Supplementary Figure A.4. We observed the similar trend with ISD prediction as well, with the significant C-index improvement of 15.3% (from 59% to 74.3%) using transfer learning.

4 Discussion and Conclusion

Pandemic outbreaks such as COVID-19 occur unexpectedly, but need immediate action due to their potential devastating consequences on human well-being. However, it is difficult to collect and curate enough medical data especially in early phases of a pandemic to develop accurate prediction models. This poses significant challenges for machine/deep learning applications, which can otherwise be immensely useful to identify individuals at risk or to aid clinical policy decisions at population level. Retrospective electronic health records and historical administrative datasets can be leveraged to estimate our best guess in such situations, which can then be fine-tuned by further learning with limited available pandemic data. This study demonstrates a use case by showing substantial improvement over 3 different COVID-19 related diagnostic and prognostic prediction tasks with pre-pandemic transfer learning. We expect that a similar modeling strategy can be useful in developing early artificial intelligence based solutions during pandemic outbreaks in the future.

References

[1] World Health Organization et al. COVID-19 weekly epidemiological update, edition 99, 6 july 2022. 2022.

[2] Riccardo M Inciardi, Laura Lupi, Gregorio Zaccone, Leonardo Italia, Michela Raffo, Daniela Tomasoni, Dario S Cani, Manuel Cerini, Davide Farina, Emanuele Gavazzi, et al. Cardiac involvement in a patient with coronavirus disease 2019 (COVID-19). *JAMA cardiology*, 5(7):819–824, 2020.

[3] Deepalakshmi Kaliyaperumal, Kumar Bhargavi, Karthikeyan Ramaraju, Krishna S Nair, Sudha Ramalingam, and Murali Alagesan. Electrocardiographic changes in COVID-19 patients: A hospital-based descriptive study. *Indian Journal of Critical Care Medicine: Peer-reviewed, Official Publication of Indian Society of Critical Care Medicine*, 26(1):43, 2022.

[4] Ajit Magadum and Raj Kishore. Cardiovascular manifestations of COVID-19 infection. *Cells*, 9(11):2508, 2020.

[5] Neda Dianati Maleki, Arash Ehteshami Afshar, and Paul W Armstrong. Use of electrocardiogram indices of myocardial ischemia for risk stratification and decision making of reperfusion strategies. *Journal of Electrocardiology*, 47(4):520–524, 2014.

[6] Omneya Attallah. An intelligent ECG-based tool for diagnosing COVID-19 via ensemble deep learning techniques. *Biosensors*, 12(5):299, 2022.

[7] Arun R Sridhar, Zih-Hua Chen, Jacob J Mayfield, Alison E Fohner, Panagiotis Arvanitis, Sarah Atkinson, Frieder Braunschweig, Neal A Chatterjee, Alessio Falasca Zamponi, Gregory Johnson, et al. Identifying risk of adverse outcomes in COVID-19 patients via artificial intelligence–powered analysis of 12-lead intake electrocardiogram. *Cardiovascular digital health journal*, 3(2):62–74, 2022.

[8] RR Van de Leur, H Bleijendaal, K Taha, T Mast, JMIH Gho, M Linschoten, B van Rees, MTHM Henkens, S Heymans, N Sturkenboom, et al. Electrocardiogram-based mortality prediction in patients with COVID-19 using machine learning. *Netherlands Heart Journal*, pages 1–7, 2022.

[9] Guosong Wu, Adam G D'Souza, Hude Quan, Danielle A Southern, Erik Youngson, Tyler Williamson, Cathy Eastwood, and Yuan Xu. Validity of icd-10 codes for COVID-19 patients with hospital admissions or ED visits in canada: a retrospective cohort study. *BMJ open*, 12(1):e057838, 2022.

[10] Weijie Sun, Sunil Vasu Kalmady, Amir Salimi, Nariman Sepehrvand, Eric Ly, Abram Hindle, Russell Greiner, and Padma Kaul. ECG for high-throughput screening of multiple diseases: Proof-of-concept using multi-diagnosis deep learning from population-based datasets. 2021 Medical Imaging Meets Neurips.

[11] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[12] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.

[13] Antônio H Ribeiro, Manoel Horta Ribeiro, Gabriela MM Paixão, Derick M Oliveira, Paulo R Gomes, Jéssica A Canazart, Milton PS Ferreira, Carl R Andersson, Peter W Macfarlane, Wagner Meira Jr, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature communications*, 11(1):1–9, 2020.

[14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[15] Chun-Nam Yu, Russell Greiner, Hsiu-Chin Lin, and Vickie Baracos. Learning patient-specific cancer survival distributions as a sequence of dependent regressors. *Advances in neural information processing systems*, 24, 2011.

Appendix

A Figure

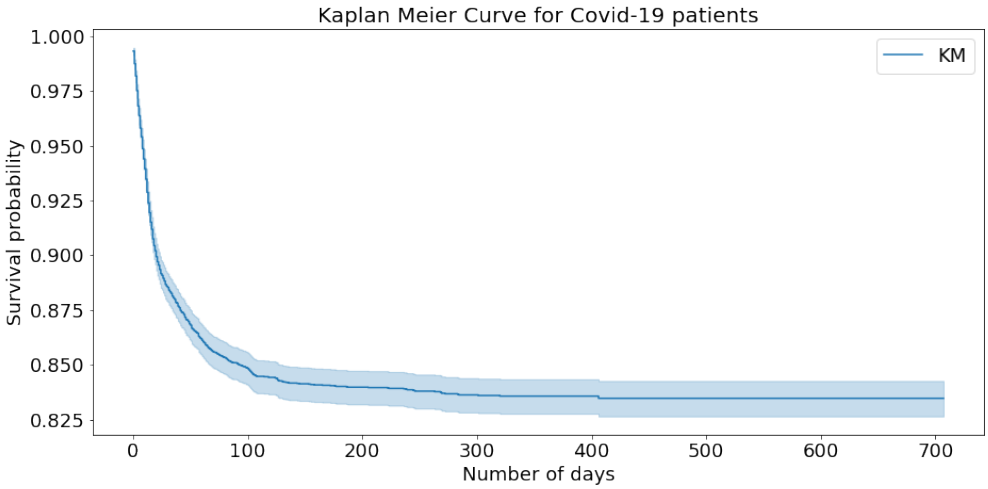

Figure A.3: Kaplan Meier Curve for patients with COVID-19 in the study cohort

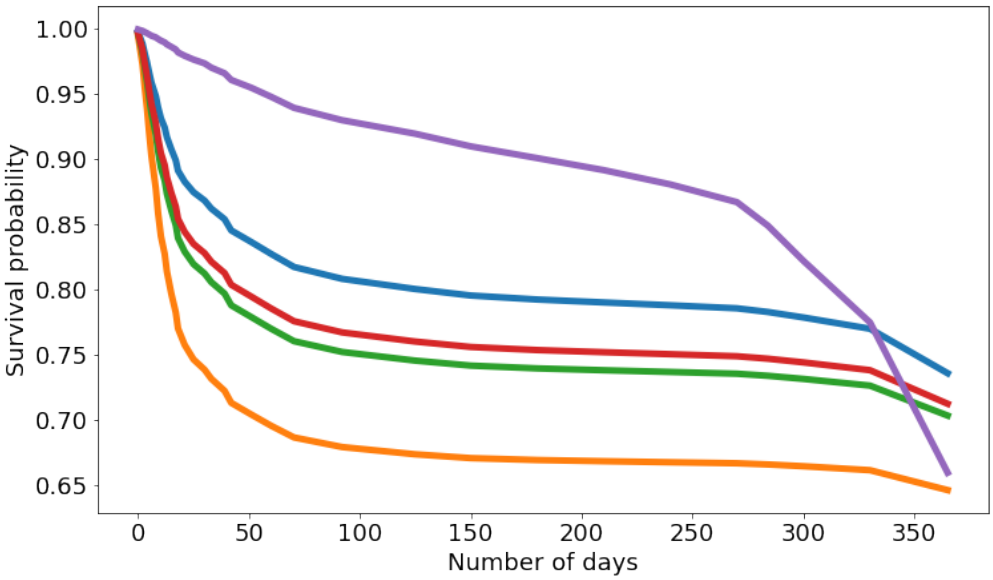

Figure A.4: Individual survival curve predictions for 5 sample COVID-19 patients

B Overview of MTLR

Consider modeling the probability of survival of patients at each of a vector of time points $\tau = [t_1, t_2, \ldots, t_m]$. In our study τ is the union of $\sqrt{N_{uncensored}}$ and month-wise in 1 years. We can set up a series of logistic regression models: For each patient, represented as $\vec{x}$,

$$P(T \geq t_i \,|\, \vec{x}) = \left(1 + \exp(\vec{\theta_i} \cdot \vec{x})\right)^{-1} \tag{1}$$

where $\vec{\theta_i}$ are the time-specific feature vectors. While the input features $\vec{x}$ stay the same for all these classification tasks, the binary labels $y_i = [T \geq t_i]$ can change depending on the threshold t_i.% We encode the survival time d of a patient as a binary sequence $y = y(d) = (y_1, y_2, \ldots, y_m)$, where $y_i = y_i(d) \in \{0, 1\}$ denotes the survival status of the patient at time t_i, so that $y_i = 0$ (no death or readmission event yet) for all i with $t_i < d$, and $(y_i = 1)$ (death) for all i with $t_i \geq d$.

Here there are $m + 1$ possible legal sequences of the form $(0, 0, \ldots, 1, 1, \ldots, 1)$, including the sequence of all 0's and the sequence of all 1's. The probability of observing the survival status sequence $y = (y_1, y_2, \ldots, y_m)$ can be represented as:

$$P_\Theta(Y{=}(y_1, y_2, \ldots, y_m) \mid \vec{x}) = \frac{\exp(\sum_{i=1}^{m} y_i \times \theta_i \cdot \vec{x})}{\sum_{k=0}^{m} \exp(f_\Theta(\vec{x}, k))}, \tag{2}$$

where $\Theta = (\theta_1, \ldots, \theta_m)$, and $f_\Theta(\vec{x}, k) = \sum_{i=k+1}^{m} (\theta_i \cdot \vec{x})$ for $0 \leq k \leq m$ is the score of the sequence with the event occurring in the interval $[t_k, t_{k+1})$ before taking the logistic transform, with the boundary case $f_\Theta(\vec{x}, k) = 0$ being the score for the sequence of all '0's. Given a dataset of n patients $\{\vec{x_r}\}$ with associated time of events $\{d_r\}$, we find the optimal parameters (for the MTLR model) (Θ^*) as

$$\Theta^* \;=\; \arg\max_{\Theta} \sum_{r=1}^{n} \left[\sum_{i=1}^{m} y_j(d_r)(\theta i \cdot \vec{x_r}) - \log \sum_{k=0}^{m} \exp f_\Theta(\vec{x_r}, k) \right] - \frac{C}{2} \sum_{j=1}^{m} \|\theta_j\|^2 \tag{3}$$

where the C (for the regularizer) is found by an internal cross-validation process. There are many details here – to insure that the survival function starts at 1.0, and decreases monotonically and smoothly until reaching 0.0 for the final time point; to deal appropriately with censored patients; to decide how many time points to consider (m); and to minimize the risk of overfitting (by regularizing), and by selecting the relevant features.

The paper by Yu et al. provides the details. C.-N. Yu, R. Greiner, H.-C. Lin, and V. Baracos. Learning patient-specific cancer survival distributions as a sequence of dependent regressors. In NIPS, 2011[15].

C Performance table

Task	COVID-19 Diagnosis		30-days mortality		Task	ISD mortality	
Transfer learning	No	Yes	No	Yes	Transfer learning	No	Yes
AUROC	60.70%	71.80%	60.10%	70.70%	C index	59.00%	74.30%
AUPRC	6.49%	13.00%	15.74%	24.83%	L1 marginal loss	2047.98	2057.65
AP	6.54%	13.10%	16.03%	25.11%	L1 hinge loss	79.29	74.91
F1	10.70%	16.40%	24.84%	30.28%			
Specificity	57.70%	79.70%	77.25%	76.41%			
Recall	58.50%	48.70%	40.69%	52.41%			
Precision	5.90%	9.83%	17.88%	21.29%			
Accuracy	57.70%	78.40%	73.27%	73.80%			
Brier score	4.16%	4.15%	9.72%	9.53%			
P+ (test)	1380	1380	145	145			
P- (test)	30377	30377	1191	1191			
P+ (all)	9090	9090	1064	1064			
P- (all)	190768	190768	7836	7836			

Table C.1

Performance of COVID-19 prediction models with and without pre-pandemic transfer learning[2]

[2]AUROC: area under the receiver operating characteristic curve; AUPRC: Area Under the Precision-recall Curve; AP: Average Precision; P+/- (test): positive/negative instances in test set; P+/- (all): positive/negative instances in all sets; ISD: individual survival distribution; C index: concordance index