Learning Deeply Enriched Representations of Temporal Data of COVID-19 Patients for Improved Mortality Prediction

Hoon Seo, Ibrohim Nosirov, and Hua Wang

Colorado School of Mines, Golden, CO 80401, USA

Abstract. An influx of COVID-19 infections puts pressure on healthcare systems and disrupts their general care routine, and efficient allocation of finite resources becomes a crucial problem. To aid logistical decision making, we propose novel semi-supervised framework to predict clinical outcomes from longitudinal records of COVID-19 patients. We use Long Short Term Memory (LSTM) architecture to learn the fixed length vectorial abstract representation summarizing the incomplete time series with varying length and uneven time intervals. In our experiments, our approach predicts mortality in more than 10 days with the blood samples collected from the 358 patients infected with COVID-19 in Wuhan, China. Our embedding framework shows 88% to 94% prediction accuracy, even if very few samples are labeled. We release our code for implementation of proposed framework.[1]

Keywords: COVID-19 · Longitudinal Learning · Semi-supervised · Enrichment

1 Introduction

Sudden increases in COVID-19 cases, such as during seasonal waves, quickly deplete the limited resources of health care systems, forcing clinicians to set criteria for distribution of scarce treatments [3]. In an earlier study, Yan et al. advocated for a machine learning mortality-prediction model to inform logistical planning [15]. The model [15] proposed in the study identifies the most predictive biomarkers for the patient's mortality using a random forest algorithm trained on a publicly available time series dataset collected from COVID-19-positive patients admitted to Tongji Hospital in China. Acknowledging the limitations of random forest models in synthesizing longitudinal data, only the final sample was used to train and test the model. While the model accomplishes the task of determining the most predictive biomarkers, it fails to capture the temporal trends, which are crucial to capturing the progression of this disease.

Conventional classification methods in time series analysis usually require sequences to be of consistent length [9], which is rarely the case when data is

[1] During review, our code is hosted in the Supplementary Information on EasyChair.

collected outside of a carefully controlled environment. Our samples were collected during an outbreak of a pandemic, a time when hospitals function out of routine and blood draws are not performed for a controlled study. As a result, the number of blood draws is different for every patient and time intervals between samples are also often irregular. The data for this specific dataset was also conducted using CRFs (case report forms), or surveys filled out by individuals rather than being transcribed directly from a test result. The resultant inputs are uneven sequences of vectors with missing information that introduce several new difficulties, particularly when the missing information includes patient outcome, the target variable for any predictive model.

One might suggest imputation as a possible solution, at least to tackle the missing data problem. Modern imputation methods rely on a framework of generative adversarial networks (GANs) [16] to successfully impute sizeable portions of high-dimensional data. Although they are efficient, even state-of-the-art techniques assume records to be missing-at-random, an assumption that risks introducing unintended bias into a model. If we suppose that the data is perfectly imputed and that we are able to coerce a classification model to recognize time as a feature, there is still variability within individual samples. Patient blood is sampled only when the clinician deems it necessary, so the number of records in a sequence, as well as the time interval between samples, is most likely different for every patient. Many time series models require the number of the records to be uniform across sequences while techniques that are reliant on time underlying time functions, such as autoencoders, fail completely when the intervals between records are non-uniform [17]. All of these difficulties are typically tackled by removing samples, a practice that leads to a smaller dataset and a less accurate model.

Most importantly, data collected in the field usually consists of labeled and unlabeled samples, and the supervised [15] or unsupervised [6, 12] learning models cannot fully utilize both samples. Therefore we propose an alternate solution in the form of a semi-supervised enrichment method to compress the incomplete time series with varying length and uneven time intervals. We outline the proposed model in Fig 1.

2 Methods

In this section, we describe the semi-supervised AutoEncoder (AE) structure to learn a fixed length vectorial representation for each patient's longitudinal records with varying length. The proposed AE is a composite model of three components. The encoder ϕ_E learns an abstract enriched representation of the input time series data. The decoder ϕ_D learns the reconstruction system to estimate the original data from the enriched representation. The predictor ϕ_P learns the prediction system to estimate the target labels from the enriched representation.

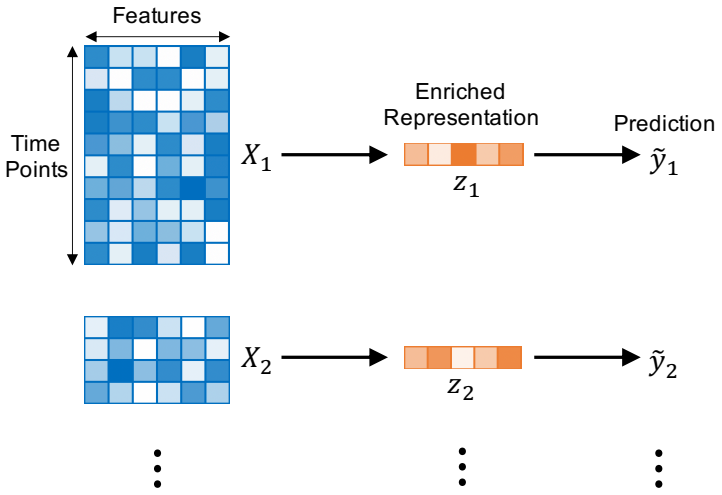

Fig. 1. The prediction via enrichment. We first compress the incomplete time series data with uneven time intervals and varying length to the fixed length vectorial representation, and then predict target labels from the enriched vectorial representation.

2.1 Notation

Dataset In the following pages, we denote a vector as a bold lower case letter, and matrix as a bold upper case letter. We describe the records of i-th patient as $\mathcal{X}_i = \{\mathbf{x}_i^s, \mathbf{X}_i, \mathbf{M}_i, \mathbf{t}_i\}$. $\mathbf{x}_i^s \in \Re^{D_s}$ is a static record such as age and admission time, and $\mathbf{X}_i = [\mathbf{x}_i^1; \mathbf{x}_i^2; \cdots; \mathbf{x}_i^{n_i}] \in \Re^{n_i \times D_l}$ are the longitudinal records collected across the n_i time points, and $\mathbf{M}_i = [\mathbf{m}_i^1; \mathbf{m}_i^2; \cdots; \mathbf{m}_i^{n_i}] \in \{1, 0\}^{n_i \times D_l}$ are the binary masks of longitudinal records $\mathbf{X}_i$, where 1 and 0 indicates the observed and unobserved entry respectively. $\mathbf{t}_i = [t_i^1; t_i^2; \cdots; t_i^{n_i}] \in \Re^{n_i}$ are the time stamps of n_i records. The target label y_i of i-th patient is given for the labeled sample in training set, such that $i \in \Omega$.

Autoencoder The encoder $\phi : \Re^{n_i \times (2D_l+1)} \mapsto \Re^{D_z}$ maps the input time series in the high dimensional space into the latent representation in the lower dimensional space, such that $\phi_E(\mathbf{X}_i, \mathbf{M}_i, \mathbf{t}_i; \theta_E) = \mathbf{z}_i \in \Re^{D_z}$. From the enriched representation $\mathbf{z}_i$, decoder $\phi_D : \Re^{D_z+1} \mapsto \Re^{D_l}$ reconstructs the original record $\phi_D(\mathbf{z}_i, t_i^j; \theta_D) = \tilde{\mathbf{x}}_i^j \in \Re^{D_l}$, for each time point t_i^j $(1 \leq j \leq n_i)$. The predictor $\phi_P : \Re^{D_z+D_s} \mapsto [0, 1]$ generates the prediction $\phi_P(\mathbf{z}_i, \mathbf{x}_i^s; \theta_P) = \tilde{y}_i$. D_s, D_l, D_z denote the dimensionality of static, longitudinal, enriched biomarkers, and $\theta_E, \theta_D, \theta_P$ denote the trainable parameters of encoder, decoder, predictor, respectively.

2.2 Encoder

We leverage an LSTM encoder to compress the time series and extract long term dependencies in the temporal trend.

Input Because the time stamp of each record is crucial to learn the temporal trend and missingness pattern of the entries may represent the state of patient, we provide the concatenation of longitudinal records, masks, and time stamps $[\mathbf{X}_i, \mathbf{M}_i, \mathbf{t}_i] = \hat{\mathbf{X}}_i = [\hat{\mathbf{x}}_i^1; \hat{\mathbf{x}}_i^2; \cdots; \hat{\mathbf{x}}_i^{n_i}] \in \Re^{n_i \times (2D_l+1)}$ to the input layer of LSTM encoder.

Output For each time step ($1 \leq j \leq n_i$), the input record $\hat{\mathbf{x}}_i^j$ of i-th patient is processed by following the LSTM architecture [18]:

$$\mathbf{k}_i^j = \sigma(\hat{\mathbf{x}}_i^j \mathbf{W}_{xk} + \mathbf{h}_i^{j-1} \mathbf{W}_{hk} + \mathbf{c}_i^{j-1} \mathbf{W}_{ck} + \mathbf{b}_k), \tag{1}$$

$$\mathbf{f}_i^j = \sigma(\hat{\mathbf{x}}_i^j \mathbf{W}_{xf} + \mathbf{h}_i^{j-1} \mathbf{W}_{hf} + \mathbf{c}_i^{j-1} \mathbf{W}_{cf} + \mathbf{b}_f), \tag{2}$$

$$\mathbf{c}_i^j = \mathbf{f}_i^j \odot \mathbf{c}_i^{j-1} + \mathbf{k}_i^j \odot \tanh(\hat{\mathbf{x}}_i^j \mathbf{W}_{xc} + \mathbf{h}_i^{j-1} \mathbf{W}_{hc} + \mathbf{b}_c), \tag{3}$$

$$\mathbf{o}_i^j = \sigma(\hat{\mathbf{x}}_i^j \mathbf{W}_{xo} + \mathbf{h}_i^{j-1} \mathbf{W}_{ho} + \mathbf{c}_i^j \mathbf{W}_{co} + \mathbf{b}_o), \tag{4}$$

$$\mathbf{h}_i^j = \mathbf{o}_i^j \odot \tanh(\mathbf{c}_i^j), \tag{5}$$

where σ is logistic sigmoid activation function and $\mathbf{k}_i^j$, $\mathbf{o}_i^j$, $\mathbf{f}_i^j$ are input, output, forget gate of j-th time step respectively. $\{\mathbf{W}_{xk}, \mathbf{W}_{hk}, \mathbf{W}_{ck}, \mathbf{W}_{xf}, \mathbf{W}_{hf}, \mathbf{W}_{cf}, \mathbf{W}_{xc}, \mathbf{W}_{hc}, \mathbf{W}_{xo}, \mathbf{W}_{ho}, \mathbf{W}_{co}\} \subset \theta_E$ are trainable weights matrix and $\{\mathbf{b}_k, \mathbf{b}_f, \mathbf{b}_c, \mathbf{b}_o\} \subset \theta_E$ are trainable bias vectors. $\mathbf{c}_i^j$ and $\mathbf{h}_i^j$ denote the cell state and hidden representation.

In order decoder and predictor to reconstruct the original record and predict the target label accurately, the enriched representation should contain enough information for reconstruction and prediction, such as temporal trend of input sequence. The goal of encoder at j-th time pont is to generate the hidden representation $\mathbf{h}_i^j$ summarizing the records from first time step to j-th time step, therefore the LSTM cell refers to the cell state $\mathbf{c}_i^j$ and reflect the past records to $\mathbf{h}_i^j$. Since the cell state $\mathbf{c}_i^j$ is guided by the input gate $\mathbf{k}_i^j$ and forget gate $\mathbf{f}_i^j$ which control how much information came from previous step should be preserved, the cell state $\mathbf{c}_i^j$ enables the hidden representation $\mathbf{h}_i^j$ to learn the long term dependencies. For example, if the differences in the records and time intervals between the time points of recent 10 consecutive records are small, the hidden representation of 10 time steps before can be preserved. The hidden representation $\mathbf{h}_i^{n_i}$ at the most recent time point is the enriched representation $\mathbf{z}_i$ of whole longitudinal records of i-th patent:

$$\mathbf{h}_i^{n_i} = \mathbf{z}_i = \phi_E(\mathbf{X}_i, \mathbf{M}_i, \mathbf{t}_i; \theta_E). \tag{6}$$

2.3 Decoder and Predictor

From the abstract representation $\mathbf{z}_i$ of longitudinal records, the decoder reconstruct the original record, and predictor predicts the target label. We use the multilayer perceptron (MLP) as the decoder and predictor, instead of additional LSTM. MLP consists of the consecutive hidden layers as follows:

$$\mathbf{h}_k = \sigma(\mathbf{W}_k \mathbf{h}_{k-1} + \mathbf{b}_k), \tag{7}$$

where $\mathbf{h}_k$ is the output of k-th hidden layer, and σ is the activation function, and $\mathbf{W}_k$, $\mathbf{b}_k$ are the trainable weights matrix and bias vector of k-th hidden layer. The output is earned by forwarding the input vector $\mathbf{h}_0$ to the last hidden layer.

To recover the original record at the specific time point, the decoder needs to know that time point. Thus the input vector of decoder is the concatenation of encoded representation $\mathbf{z}_i$ and time stamp t_i^j, such that $\mathbf{h}_0 = [\mathbf{z}_i, t_i^j] \in \Re^{D_z+1}$. For the predictor, the static information such as age may be crucial to predict target label. Therefore we provide the abstract representation of time series with static information to the predictor, such that $\mathbf{h}_0 = [\mathbf{z}_i, \mathbf{x}_i^s] \in \Re^{D_z+D_s}$. By forwarding each input of decoder and predictor to their own MLPs, we earn the reconstructed record $\tilde{\mathbf{x}}_i^j$ and prediction $\tilde{y}_i$ on target label:

$$\phi_D(\phi_E(\mathbf{X}_i, \mathbf{M}_i, \mathbf{t}_i; \theta_E), t_i^j; \theta_D) = \tilde{\mathbf{x}}_i^j, \tag{8}$$

$$\phi_P(\phi_E(\mathbf{X}_i, \mathbf{M}_i, \mathbf{t}_i; \theta_E), \mathbf{x}_i^s; \theta_P) = \tilde{\mathbf{y}}_i, \tag{9}$$

and we have the stack of reconstructed records of i-th patient:

$$\tilde{\mathbf{X}}_i = [\tilde{\mathbf{x}}_i^1; \tilde{\mathbf{x}}_i^2; \cdots; \tilde{\mathbf{x}}_i^{n_i}]. \tag{10}$$

The composite model of encoder, decoder, and predictor is illustrated in Fig. 2

2.4 Loss Function

AE is asked to accomplish the two tasks - reconstructing original records and predicting target label by minimizing:

$$\min_{\theta_E, \theta_D, \theta_P} \mathcal{L}_{total} = \min_{\theta_E, \theta_D, \theta_P} (\gamma_1 \mathcal{L}_{reconstruct} + \gamma_2 \mathcal{L}_{predict}), \tag{11}$$

where γ_1 and γ_2 are the hyperparameters to adjust the impact of each loss. The reconstruction loss is defined as Mean Squared Error (MSE):

$$\mathcal{L}_{reconstruct} = \frac{\left\| (\tilde{\mathbf{X}}_i - \mathbf{X}_i) \odot \mathbf{M}_i \right\|_F^2}{|\mathbf{M}_i|}, \tag{12}$$

where squared Frobenious norm $\| \cdot \|_F^2$ is defined as the summation of all the entries squared. Since the semi-supervised learning model needs to utilize both labeled and unlabeled data, the loss function should be well defined respect to

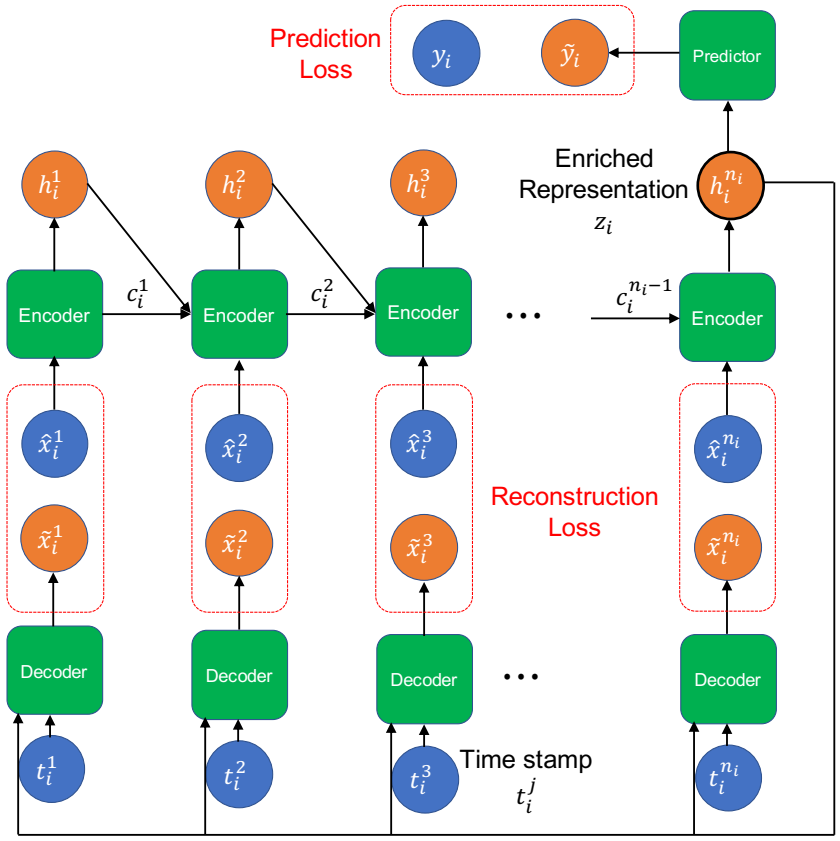

Fig. 2. An illustration of proposed semi-supervised autoencoder. The encoder consists of LSTM cell, which processes input record $\hat{\mathbf{x}}_i^j$, generates the hidden representation $\mathbf{h}_i^j$ while conveying the cell state $\mathbf{c}_i^j$. The enriched representation of the whole time series is the hidden representation at the most recent time step.

labeled $i \in \Omega$ and unlabeled $i \notin \Omega$ data both. Therefore the prediction loss respect to the target label is defined as:

$$\mathcal{L}_{predict} = \left\{ \begin{matrix} \|\tilde{y}_i - y_i\|_F^2, & \text{for } i \in \Omega \\ 0, & \text{for } i \notin \Omega \end{matrix} \right\}. \tag{13}$$

The high capacity unsupervised AE with only reconstruction error may suffer from the tendency to learn the trivial identity mapping and memorize the input [12]. The addition of prediction loss can prevent the memorization, because the memorization is not useful to predict the target label.

3 Experiment

3.1 Dataset

We obtain the blood sample records, basic demographic information, and associated mortality outcomes of 375 patients collected throughout their stay in hospital from Tongji Hospital between 10th January and 24th February 2020 following the previous research [15]. We discard 17 samples whose no time stamp is recorded. Among the remaining 358 samples, 192 patients survived (labeled by 1) and 166 patients died (labeled by 0). The order of samples is randomly shuffled to prevent the bias from the order. The missing entries are initialized as -1. Each feature of dataset is normalized by min-max scaling to the range [0, 1].

3.2 Hyperparameters

We find the best hyperparameters of our model by the grid search; LSTM units in {64, 100, 128}, γ_1 and γ_2 in $\{1e-1,\ 1e-2,\ 1e-3\}$, and learning rate of optimizer in $\{1e-2,\ 1e-3,\ 1e-4,\ 1e-5\}$. We construct the decoder and predictor with 3 hidden layers, and the number of hidden layers and units of each layer does not impact the performance significantly. We use hyperbolic tangent activation function for all layers, except logistic sigmoid function at the output layer of predictor, and leaky Rectified Linear Unit (leaky ReLU, alpha = 0.1) at the input layer of decoder. We note that the leaky ReLU have been largely improved the reconstruction performance of decoder, as we presume that time stamp is the most important input for the decoder and this can be emphasized more with leaky ReLU activation function in a range $(-\infty, \infty)$, than the other activation functions in a smaller range. The mini batch size is set to 1. We do not use any regularization or dropout technique, as they are found to degrade the performance. We minimize the loss function in Eq. (11) with Adam optimizer [4].

3.3 Mortality Classification Task

To evaluate the prediction performance of the proposed model, we consider the classification problem to predict the mortality of COVID-19 patients in advance more than 10 days. The output of predictors are rounded up to the binary. We compare the accuracy, precision, recall, and F_1-score of our model (Semi-supervised AutoEncoder, SAE) to those metrics of the other baseline supervised learning models. The pure LSTM model is introduced for the ablation study, and it consists of LSTM encoder and MLP predictor same as our model, but the decoder is removed. We conduct the k-fold cross validation by splitting the dataset samples into k sub-groups, and calculate the average and standard deviation of the four metrics over the k sub-groups. The baseline supervised learning models are trained on the training set, while our semi-supervised learning model is trained on training and test set both. The blood sample records at all the time points are given to the time series models, LSTM and SAE, while the blood sample record only at the last time point is given to the other baseline models.

The demographic information, age, gender, admission time, discharge time, are provided as the static record to all prediction models.

Table 1. The prediction results of SAE and the baseline models from k-fold cross validation. The best prediction is denoted as bold font.

Test Set Size	Model	Accuracy	Precision	Recall	F_1-score
20%	SAE	**94.14±2.31**	**93.35±2.45**	**93.15±2.84**	**93.48±1.21**
	LSTM	92.56±3.24	91.11±1.45	92.45±8.12	91.43±5.40
	MLP	80.63±11.52	76.58±10.94	83.80±11.97	80.02±11.14
	RF	83.25±11.89	78.03±11.15	88.77±12.68	83.05±11.86
	RR	79.32±11.33	76.76±10.97	79.54±11.36	78.12±11.16
	SVM	79.32±11.33	77.55±11.08	78.12±11.16	77.83±11.11
25%	SAE	**92.47±3.43**	**91.84±8.41**	**91.97±2.34**	**91.63±3.94**
	LSTM	91.91±3.69	91.38±8.94	91.62±1.89	91.17±3.98
	MLP	81.90±11.70	79.25±11.32	82.92±11.85	81.05±11.58
	RF	83.30±11.9	78.94±11.28	87.45±12.49	82.98±11.85
	RR	80.50±11.50	81.13±11.16	76.14±10.88	78.56±11.22
	SVM	80.50±8.12	80.11±11.45	77.65±11.09	78.86±11.27
75%	SAE	**89.48±1.59**	**88.59±2.27**	88.72±6.09	**88.46±2.1**
	LSTM	87.91±1.07	87.41±3.22	86.07±4.59	86.57±1.4
	MLP	80.09±11.44	74.94±10.71	86.86±12.41	80.46±11.50
	RF	88.52±12.65	85.49±12.21	**91.32±13.0**	88.31±12.62
	RR	79.03±11.29	75.54±10.79	82.41±11.77	78.82±11.26
	SVM	77.98±11.11	75.06±10.7	80.18±11.45	77.54±11.08
80%	SAE	**88.06±1.07**	**86.98±1.82**	**87.03±3.29**	**86.95±1.44**
	LSTM	84.07±5.55	86.55±2.91	77.60±16.00	80.66±9.94
	MLP	79.46±11.35	72.80±10.40	82.19±11.17	77.21±11.03
	RF	51.15±4.09	64.18±14.00	30.00±25.83	59.81±16.37
	RR	78.14±11.16	70.78±10.11	82.19±11.17	76.06±10.87
	SVM	74.16±10.79	66.22±9.46	79.03±11.3	72.06±10.29

The experimental results in Table 3.3 show that our model outperforms the other baseline models for all the different size of test sets. The performance improvement's margins of our model respect to the baseline models are larger when the smaller size of training set (which means the larger size of test set) is given. We suppose this is because our model is able to learn from the unlabeled samples, while the other supervised learning models are able to learn from only the labeled samples. The unlabeled samples are also useful resources for our model to improve the enrichment capability, which can help the prediction from the enriched representation.

3.4 Biomarker Identification

We identifies which biomarkers of blood samples are the risk factors for the mortality of COVID-19 patients. Despite the high performance of artificial neural

networks, the outputs of artificial neural networks are notoriously difficult to be interpreted. To identify which biomarkers largely affect to the predictions, we add the perturbations to the input time series and observe the prediction changes. We try to perturb each feature of time series in the range of the distribution of the feature.

For m-th feature ($1 \leq m \leq D_l$) and i-th patient, we sample the column vector of perturbation $\mathbf{p}_{i,m} \in \Re^{n_i}$ from the normal distribution $\mathcal{N}(0, \sigma_m^2)$ with zero mean and the same standard deviation as the distribution of m-th feature across all the time and patients, as follows:

$$\begin{aligned} N &= \sum_{i=1}^{n}\sum_{j=1}^{n_i} m_i^j,\ \mu_m = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i} m_i^j * x_i^j, \\ \sigma_m^2 &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i} m_i^j (x_i^j - \mu_m)^2. \end{aligned} \tag{14}$$

Then the time series whose m-th feature is perturbed and it's prediction change is:

$$\begin{aligned} &\mathbf{X}_i' = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \cdots, \mathbf{x}_{i,m} + \mathbf{p}_{i,m}, \cdots, \mathbf{x}_{i,D_l}], \\ &\Delta\tilde{\mathbf{y}}_i = \|\phi_P(\phi_E(\mathbf{X}_i, \mathbf{M}_i, \mathbf{t}_i; \theta_E), \mathbf{x}_i^s; \theta_P) - \phi_P(\phi_E(\mathbf{X}_i', \mathbf{M}_i, \mathbf{t}_i; \theta_E), \mathbf{x}_i^s; \theta_P)\|, \end{aligned} \tag{15}$$

where $\mathbf{x}_{i,m} \in \Re^{n_i}$ is the column vector of feature observations of i-th patient across all the time points. Then the relative importance of m-th feature is the average of prediction changes across all the patients: $\frac{1}{n}\sum_{i=1}^{n} \Delta\tilde{\mathbf{y}}_i$. Finally, we plot top 15 most important features in Fig. 3.

Some of the identified biomarkers are consistent with the previous medical findings. For example, lactic dehydrogenase (LDH), lymphocyte and high-sensitivity C-reactive protein (hs-CRP) are the top 3 biomarkers relevant with the mortality of COVID-19 patients, identified by the XGBoost model [15] and previous medical researches [5, 10, 14]. To be specific, the increase of LDH indicates tissue or cell destruction and this is the strong sign of tissue or cell damage [15]. The activity of idiopathic pulmonary fibrosis can be detected by Serum LDH [5]. The hs-CRP is the risk factor for the continuous inflammation [1] and poor prognosis in acute respiratory distress syndrome [5, 11]. The lymphocyte is the common risk factor of COVID-19 patients [2], and lymphocyte has relation with the decrease in CD4 and CD8 T cells [8]. Albumin have been found to be independently associated with mortality, at the Cox regression analysis [13]. The basophil count is known to be the risk factor of mortality and organ injury in COVID-19 patients [7]. The identified biomarkers are supported by the previous studies, and add the value on our semi-supervised prediction model and perturbation based analysis.

4 Conclusion

We propose semi-supervised enrichment methods based on an LSTM autoencoder. The enriched representation in a fixed length vectorial format summarizes

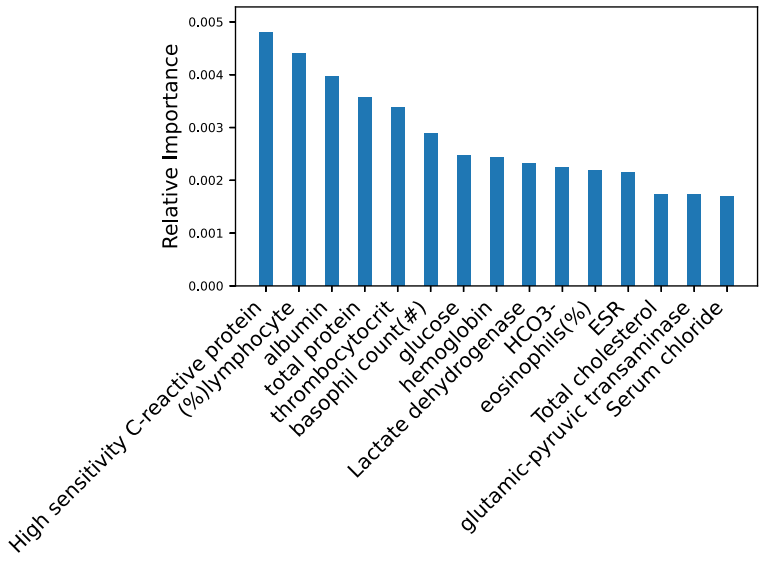

Fig. 3. Top 15 important biomarkers identified by the proposed model.

time series data with uneven time intervals and possible missing entries, allowing the decoder to explicitly utilize the uneven time intervals to reconstruct the original data. Our model is able to learn from both labeled or unlabeled samples both jointly, producing state-of-the-art prediction performance on mortality prediction. Armed with the trained model, our model can readily predict the mortality of a new patient free from the cold start problem. Our model is purely data-driven, which means the prediction performance can be further improved with additional data. Since there is no assumption or limitation in predictor structure or target labels, the other models stemming from our approach will be flexible to solve the other problems, such as multi-label classification or regression problem.

References

1. Bajwa, E.K., Khan, U.A., Januzzi, J.L., Gong, M.N., Thompson, B.T., Christiani, D.C.: Plasma c-reactive protein levels are associated with improved outcome in ards. Chest **136**(2), 471–480 (2009)
2. Chan, J.F.W., Yuan, S., Kok, K.H., To, K.K.W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C.Y., Poon, R.W.S., et al.: A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The Lancet **395**(10223), 514–523 (2020)
3. for Disease Control, C., (CDC), P., et al.: Strategies for optimizing the supply of n95 respirators. 2020 apri 3 https://www. cdc. gov/coronavirus/2019-ncov/hcp/respiratorsstrategy/index. html. CDC_AA_refVal= https% 3A% 2F% 2Fwww. cdc. gov% 2Fcoro navirus% 2F2019-ncov% 2Fhcp% 2Frespiratorsstrategy% 2Fcrisis-alternate-strategies
4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
5. Kishaba, T., Tamaki, H., Shimaoka, Y., Fukuyama, H., Yamashiro, S.: Staging of acute exacerbation in patients with idiopathic pulmonary fibrosis. Lung **192**(1), 141–149 (2014)
6. Längkvist, M., Karlsson, L., Loutfi, A.: A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recognition Letters **42**, 11–24 (2014)
7. Li, D., Chen, Y., Liu, H., Jia, Y., Li, F., Wang, W., Wu, J., Wan, Z., Cao, Y., Zeng, R.: Immune dysfunction leads to mortality and organ injury in patients with covid-19 in china: insights from ers-covid-19 study. Signal Transduction and Targeted Therapy **5**(1), 1–3 (2020)
8. Liu, J., Li, S., Liu, J., Liang, B., Wang, X., Wang, H., Li, W., Tong, Q., Yi, J., Zhao, L., et al.: Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of sars-cov-2 infected patients. EBioMedicine p. 102763 (2020)
9. Lu, L., Wang, H., Yao, X., Risacher, S., Saykin, A., Shen, L.: Predicting progressions of cognitive outcomes via high-order multi-modal multi-task feature learning. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 545–548. IEEE (2018)
10. Ridker, P.M., Danielson, E., Fonseca, F.A., Genest, J., Gotto Jr, A.M., Kastelein, J.J., Koenig, W., Libby, P., Lorenzatti, A.J., MacFadyen, J.G., et al.: Rosuvastatin to prevent vascular events in men and women with elevated c-reactive protein. New England journal of medicine **359**(21), 2195–2207 (2008)
11. Sharma, S.K., Gupta, A., Biswas, A., Sharma, A., Malhotra, A., Prasad, K., Vishnubhatla, S., Ajmani, S., Mishra, H., Soneja, M., et al.: Aetiology, outcomes & predictors of mortality in acute respiratory distress syndrome from a tertiary care centre in north india. The Indian journal of medical research **143**(6), 782 (2016)
12. Srivastava, N., Mansimov, E., Salakhudinov, R.: Unsupervised learning of video representations using lstms. In: International conference on machine learning. pp. 843–852 (2015)
13. Violi, F., Cangemi, R., Romiti, G.F., Ceccarelli, G., Oliva, A., Alessandri, F., Pirro, M., Pignatelli, P., Lichtner, M., Carraro, A., et al.: Is albumin predictor of mortality in covid-19? Antioxidants and Redox Signaling (ja) (2020)
14. Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., et al.: Clinical characteristics of 138 hospitalized patients with 2019

novel coronavirus–infected pneumonia in wuhan, china. Jama **323**(11), 1061–1069 (2020)
15. Yan, L., Zhang, H.T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., Sun, C., Tang, X., Jing, L., Zhang, M., et al.: An interpretable mortality prediction model for covid-19 patients. Nature Machine Intelligence pp. 1–6 (2020)
16. Yoon, J., Jordon, J., Van Der Schaar, M.: Gain: Missing data imputation using generative adversarial nets. arXiv preprint arXiv:1806.02920 (2018)
17. Yu, W., Zeng, G., Luo, P., Zhuang, F., He, Q., Shi, Z.: Embedding with autoencoder regularization. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 208–223. Springer (2013)
18. Yu, Y., Si, X., Hu, C., Zhang, J.: A review of recurrent neural networks: Lstm cells and network architectures. Neural computation **31**(7), 1235–1270 (2019)