A Deep Learning Approach to Handling Temporal Variation in Chronic Obstructive Pulmonary Disease Progression

Chunlei Tang
Brigham and Women's Hospital
Harvard Medical School
Boston, MA, USA
ctang5@partners.org

Joseph M. Plasek
Department of Biomedical Informatics, School of Medicine
University of Utah
Salt Lake City, UT, USA
jplasek@partners.org

Haohan Zhang
Shanghai Key Laboratory of Data Science, School of Computer Science
Fudan University
Shanghai, CHN
14300240009@fudan.edu.cn

Yun Xiong
Shanghai Key Laboratory of Data Science, School of Computer Science
Fudan University
Shanghai, CHN
yunx@fudan.edu.cn

David W. Bates
Brigham and Women's Hospital
Harvard Medical School
Boston, MA, USA
dbates@partners.org

Li Zhou
Brigham and Women's Hospital
Harvard Medical School
Boston, MA, USA
lzhou@bwh.harvard.edu

***Abstract*—Chronic Obstructive Pulmonary Disease (COPD) is a leading cause of mortality in the United States. Representing COPD progression using temporal graphs may offer critical clinical insights. Long-Short Term Memory units in recurrent neural networks can process data with constant elapsed times between consecutive elements of a sequence but cannot handle irregular time intervals (i.e., segments with unequal-time). In this study, we propose a four-layer deep learning model that utilizes a specially configured recurrent neural network to capture irregular time lapse segments. Experiments on a corpus of COPD patients' clinical notes compared to baseline algorithms showed that our model improved interpretability as well as the accuracy of estimating COPD progression.**

***Keywords*—"pulmonary disease, chronic obstructive," neural network, disease progression, medical informatics, machine learning**

I. Introduction

Chronic Obstructive Pulmonary Disease (COPD) is the third leading cause of mortality in the United States, affecting an estimated 14.7 million diagnosed patients [1]. According to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria [2], COPD can take well over ten years to evolve from Stage I (mild) to Stage IV (very severe). Mild COPD symptoms include shortness of breath, chronic cough, and fatigue whereas severe COPD can result in heart failure [3]. Unfortunately, there is currently no cure for COPD. The purpose of treatments for COPD is to lower a patient's risk of disease progression and death.

Management of COPD consists of an ongoing process of monitoring and assessing a patient's symptoms and conditions along with interventions. Commonly used pulmonary function tests for monitoring progression include: forced expiratory volume in one second (FEV1), forced vital capacity (FVC), the FEV1/FVC ratio, and slow vital capacity (SVC). Radiology examinations (e.g., chest X-ray, cardiac radiography) are often conducted for diagnosis and monitoring purposes. Critical information for classifying disease severity and predicting disease progression is usually embedded in clinical notes and radiology reports that contain interpretation of test results and clinical findings. Extracting this critical information from large-scale electronic health records (EHR) requires the development of data mining methods and computational methods [4-6].

COPD disease progression may span a decade or more (Fig. 1). Classifying a corpus of free-text clinical notes to a COPD stage, as defined at the 2001 NHLBI/WHO GOLD Workshop [7] (Fig. 2), is challenging due to the following cohort characteristics:

- *Irregular Visits:* The creation of a patient's clinical notes depends on patient's frequency of visiting affiliated medical facilities to receive care. Therefore, the temporal granularity of a patient's record may vary significantly over different time periods.
- *Incomplete Records:* EHRs for each patient may not be complete as patients may visit medical facilities outside of a health care provider network; thus, in some cases, clinical data may not be available for the entire progression of COPD.

The work was partially funded by the Brigham Care Redesign Incubator and Startup Program (BCRISP) program, the National Natural Science Foundation of China Projects No.91546105, No.U1636207, and the Shanghai Science and Technology Development Fund No.16JC1400801, No.17511105502, No.17511101702, No.16511102204.

978-1-5386-5488-0/18/$31.00 ©2018 IEEE

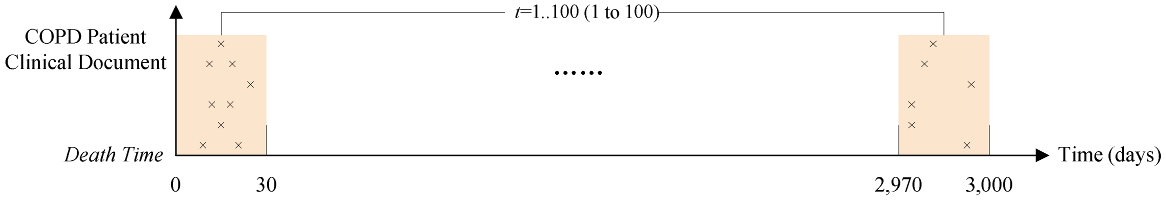


Fig. 1. An example of using the disjoint 30-day time window as a constant time segment to partition COPD disease progression as 100 time segments, in which a cross represents a clinical document falling within the window.

- *Disease Progression Heterogeneity:* Different patients may have different disease progressions. For example, COPD patients who manifest a lung infection often progress more rapidly to a more severe stage, whereas COPD patient who quit smoking often remain stable longer. There is no natural alignment between different patients as progression rates vary.
- *Discrete Observations:* Although the disease progression is a continuous-time process, the patient is only observed at discrete time points with varied intervals (e.g., 1 day for an office visit, a few days for a hospitalization).

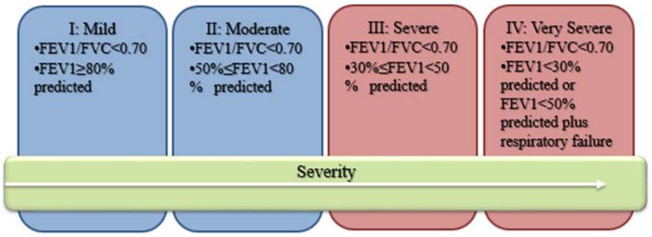


Fig. 2. Four main stages of COPD as identified in the 2001 NHLBI/WHO GOLD workshop.

When using a constant, or regular disjoint time window to model COPD progress information, a longer time window segment (e.g., 90 days) can be problematic as the data within that window may fall within multiple different stages of COPD. Therefore, a shorter time window (e.g., a disjoint 30-day window) is often used in temporal segmentation methods [4] to map a specific clinical document to a COPD stage. Nevertheless, these constant disjoint window time segments cannot show the temporal autocorrelation from the dynamics arising in the data.

One potential approach to capture the underlying structure of sequential data is to utilize Recurrent Neural Networks (RNNs) [8], particularly, Long-Short Term Memory (LSTM), which is a gated variant of RNNs that has solved the vanishing and exploding gradient problems by handling long-term event dependencies. This approach has recently been applied to medical informatics applications, with promising results [9, 10]. However, a standard LSTM setup cannot deal with irregular time intervals [5], and prior studies [5, 10] that modified the LSTM layer for accepting irregular time interval segments are limited by the need to formulate a continuous time hypothesis.

Unlike other approaches, we aimed to capture such time irregularities by incorporating COPD conditions learned from clinical notes. The possible vital periods might signify the stage transitions as elicited in the prior medical knowledge of the ground truth progression stages (e.g., the key indicators). The specific aim of this study was to establish the feasibility of utilizing deep learning to model irregular time segments and predict COPD progression.

II. Methods

We developed a four-layer deep learning model using RNNs to adjust time interval settings automatically and to capture an irregular time lapse of temporal segments (Fig. 3). The components of the model include: 1) a pre-processing and word embedding layer to prepare the data, 2) a LSTM layer to predict death date, and 3) a flatten and dense layer combination to capture the irregular time lapse of segments. Our model was implemented in Keras (version 2.2.0) – an open source neural network library written in Python (version 3.7.0).

We define the notations that we will use throughout the rest of the paper in Table 1.

TABLE I. Meaning of Notation

Notations	*Interpretation*
V	total number of words in our vocabulary of word embeddings
v	dimension of word embeddings in vector space
T	maximum words in each input document used for word embeddings
D	total number of samples; each sample formed by COPD notes within one day
B	Initial quantity of regular time interval used in LSTM
$N=D/B$	total number of batches
L	number of hidden units in the LSTM cell

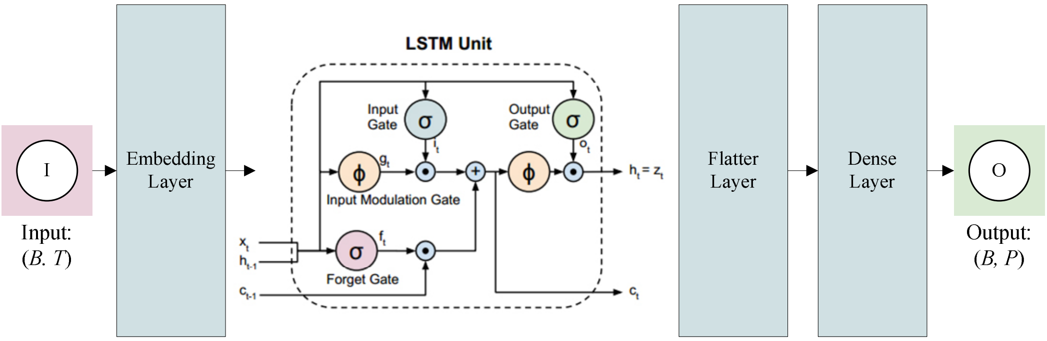


Fig. 3. An illustration of the proposed model that includes an embedding layer, long short term memory (LSTM) layer, flatten layer, and dense layer. See Table 1 and Equations (1) to (6).

A. Data description

The data used in this study consists of three types of free-text clinical notes (i.e., pulmonary notes, radiology reports, cardiology reports), which were extracted from the Research Patient Data Registry at Partners Healthcare, an integrated healthcare delivery network located in the greater Boston area of Massachusetts. We used International Classification of Diseases - Version 10 (ICD-10-CM) codes to identify COPD patients in the registry. We utilized laboratory results corresponding to the GOLD [7] and Body mass index, airflow Obstruction, Dyspnea, and Exercise capacity (BODE) [11] criteria. We retrieved patients' death dates from Massachusetts Death Certificate files. A cohort of 15,500 COPD patients who both received care at any Partners Healthcare facility and died between 2011 and 2017 was extracted. This study was approved by Partners Institutional Review Board (IRB).

Clinical notes from all three domains that were merged into a single corpus using a heuristic merger that inserts notes from each domain into the appropriate chronological place in the corpus. This resulted in each sample being equivalent to a text file containing one day of clinical notes from all three domains.

- *Pulmonary notes:* We extracted physician's interpretation of patients' lung function from pulmonary notes. Each pulmonary note contains indicators for measuring the air movement in and out of the lungs during respiratory maneuvers (e.g., FVC, FEV1, the FEV1/FVC ratio), as well as a *PHYSICIAN INTERPRETATION* section. The physician interpretation involves: (1) obstructed or not? (i.e., the FEV1/FVC ratio where <=0.70 is obstructed and >0.75 is regarded as normal); (2) any restrictive component or not? (i.e., is the FVC less than the lower limit of normal?); and (3) reduced in true restriction or normal in pseudo-restriction (i.e., is the total lung capacity known?). These indicators can be useful in evaluating the nature and severity of respiratory disease but must be interpreted in the context of the patient history and examination. Other features used from these notes include: patient ID, date of test, and last test date. A total of 78,489 pulmonary notes corresponding to 2,431 unique patients were extracted. The average time span of a patient for the pulmonary data source was 724.4 days, with a max span of 3,003 days.
- *Radiology reports:* We extracted chest X-ray radiology reports and focused on two main sections of each report: FINDINGS and IMPRESSION. We extracted and examined the free-text content corresponding to each section. Other features used from these reports include: patient ID, date of test, and last test date. In our cohort, we had 1,893,498 radiology reports corresponding to 13,414 unique patients. The average time span of a patient using the radiology data source was 843.8 days, with a max span of 2,469 days.
- *Cardiology reports:* Prior research has demonstrated that there is a relationship between COPD and cardiovascular morbidity [2]. We utilized abnormal electrocardiogram reports, and their corresponding patient ID, date of test, and last test date. In our cohort, we had 1,029,363 cardiology reports for 13,918 patients. The average time span of a patient using the cardiology data source was 740.8 days, with a max span of 2,459 days.

To understand clinical important of data correlations across the three domains, we consulted medical experts in COPD regarding the priority of the three types of clinical notes. While these experts didn't suggest a hierarchy of priority, they agreed that, in general, the three types of clinical documents might provide different perspectives of disease indicators describing a COPD patients' condition. For example, a physician's interpretation of pulmonary notes is essential for the diagnosis of COPD, which is conducted by evaluating the core functioning of the lungs. Although COPD cannot be diagnosed with a chest X-ray alone, it can help physicians evaluate shortness of breath, help support a diagnosis of COPD (to rule out lung cancer, pneumonia, tuberculosis, or other potential infections), and to detect advanced emphysema (e.g., a focal bullae or abnormal air collection within the lung). Cardiac radiography may be sensitive in the more severe stages of COPD compared to earlier stages.

B. Dimensionality of Each Layer

The input of our model is (*B*, *T*). We first flattened the 3-dimensional output matrix to a 2-dimensional vector using the flatten layer, and then reduced dimensionality via the dense layer. We then conducted vector normalization via the lambda function using a sigmoid activation function. Next, iterative learning occurred along the descending direction of gradient descent via the loss function. Output of four layers is, respectively, (*B*, *T*, *v*), (*B*, *T*, *L*), (*B*, *T*×*L*), and (*B*, *P*). We can get a value to estimate a patient's mortality if we specify *P*=1 as the output of the dense layer.

C. Pre-processing and word embeddings

A one-hot encoding allows for a more expressive representation of categorical data. We pre-processed the input data (i.e., samples) to create one-hot encodings of a given regular time interval B. Each sample is a free-text document composed by combining all patients' notes (these notes may belong to different note types) that correspond to the same time interval. Next, we utilized Keras padding to remove excess data unrelated to COPD in order to make all input samples the same length.

A word embedding is a class of unsupervised learning approaches for representing words and documents using a dense vector representation. The position of a word within the vector space is learned from text and is based on the words that surround the word when it is used. The position of a word in the learned vector space is referred to as its embedding. One popular algorithm for learning word embeddings from text is Word2vec [12]. Word embeddings are an improvement over traditional bag-of-word encoding schemes where large sparse vectors are used to represent each word or to score each word within a vector to represent an entire vocabulary. Traditional representations are sparse because the vocabularies tend to be vast and a given word or document would be represented by a large vector comprised mostly of zeroes. Instead, in an embedding, words are represented by dense vectors where a vector represents the projection of the word in continuous vector space.

Keras offers an embedding layer as a hidden layer, which can be used in neural networks for processing textual data. To use this embedding layer, Keras requires that the input data be integer encoded, so that each word is represented by a unique integer. The embedding layer is initialized with random weights and will learn an embedding for all of the words in the training dataset. Keras requires specification of the parameters V, v, and T. Based on a preliminary analysis of the length and focus of COPD notes, we defined an embedding layer with a vocabulary of 10,000 (e.g. integer encoded words from 0 to 9,999, inclusive), a vector space of 64 dimensions in which words will be embedded, and input documents that have 1,000 words each.

D. Long Short-Term Memory Unit

A recurrent neural network is a class of deep neural networks where the connections between hidden units form a directed cycle. This allows the RNN to keep prior information for hidden states in internal memory. This is why RNNs are preferred for applications where the system needs to store and update the contextual information [13]. Approaches such as Hidden Markov Models (HMM) have also been used for similar purposes; however, there are distinctive properties of RNNs that differentiates them from HMMs. For example, RNNs do not make the Markov property assumption, and they can process variable length sequences. Information on past inputs can be kept in the memory without any limitation to the amount of time that has past. However, optimization for long-term dependencies is not always possible due to the vanishing and exploding gradient problems where the value of the gradient becomes too small and too large, respectively. To be able to incorporate long-term dependencies without violating the optimization process, variants of RNNs have been proposed. One of the popular variants is Long Short-Term Memory (LSTM) which is capable of handling long-term dependencies via a gated structure [14].

A standard LSTM unit comprises of forget gates, input gates, output gates, and a memory cell, but the architecture has the implicit assumption of being uniformly distributed across the elapsed time of a sequence. Detailed mathematical expressions of the LSTM used are given below, in which (1) to (6) are the input gate, forget gate, output gate, input modulation gate, current memory, and current hidden state, respectively, corresponding to each step in Fig 1. The output of the LSTM Layer, namely reshaped 3d LSTM matrices, is intermediate results from our model. Each LSTM matrix is the output from one batch of the period.

$$i_t := \text{sigmoid}\left(W_{h_i} \times h_{t-1} + W_{x_i} \times x_t + b_i\right) \tag{1}$$

$$f_t := \text{sigmoid}\left(W_{h_f} \times h_{t-1} + W_{x_f} \times x_t + b_f\right) \tag{2}$$

$$o_t := \text{sigmoid}\left(W_{h_o} \times h_{t-1} + W_{x_o} \times x_t + b_o\right) \tag{3}$$

$$g_t := \tanh\left(W_{h_g} \times h_{t-1} + W_{x_g} \times x_t + b_g\right) \tag{4}$$

$$c_t := \left(f_t \cdot c_{t-1}\right) + \left(i_t \cdot g_t\right) \tag{5}$$

$$h_t := o_t \cdot \tanh c_t \tag{6}$$

E. Capturing of Time Lapse Segments

In order to capture irregular time lapse segments, we used a flatten layer to facilitate the unfolding process followed by a dense layer to combine the time segments into a fully-connected network. We used a sigmoid activation function to output a sequence consisting of 0 and 1 as the irregular time lapse segments for each LSTM matrix (whose length is 1). In the {0,1}-sequence, we set two or more consecutive zeros or ones as a time segment. For example, the sequence of {00010000001010111} has two irregular time segments as [00010000] and [001010111].

Pseudocode is presented below.

```
Input: reshaped LSTM Matrices $\alpha \in \mathbb{R}^{N \times B \times (T \times L)}$
Output: $bound$
Given $\alpha_i, i \in [0, \cdots, N-1]$
    Initialize variables: $count = 0, bound = iB || (i+1)B$
    while $\forall \alpha_{jk}$ $(j \in [i-1, \cdots, 0],\ k \in [B-1, \cdots, 0]) || (j \in [i, \cdots, P-1],\ k \in [0, \cdots, B-1])$ do
        $\boldsymbol{Sim} \in \mathbb{R}^B$ is Cosine Distance between $\alpha_{jk}$ and $\alpha_{ik'}, k' \in [0, \cdots, B-1])$
        if the number of invalid scalars that is greater than a threshold distance in $\boldsymbol{Sim}$ exceeds $B/2$ then
            $count = count + 1;$
            if $count > B/3$ then BREAK;
        else $count = 0;$
            update $bound$ to current day
    return $bound$
```

III. Model Customization and Evaluation

A. Model Customization

We applied our model to real-world clinical notes/reports of COPD patients. It is particularly interesting for testing our model in this clinical domain because COPD has a prolonged progression path.

Our model partitions COPD disease progression through similarity of text description in clinical notes. Therefore, the initial data observation (e.g., data distribution, data correlation) for each patient has a great impact on the interpretability of our model. We conducted a simple *k*-means clustering (*k*=10) to help understand the data distribution of the three types of clinical notes. We segmented the time dimension (a total of 3,000 days) into disjoint 30-day windows to calculate the number of notes falling within each time window (Fig. 1).

B. Model Evaluation

We evaluated our model using standard performance metrics and compared these metrics against baseline classifiers.

1) LSTM Prediction Accuracy

LSTM units as a variant of recurrent neural networks are well-suited to do classification and prediction given time lags of unknown size and duration between events. In our model, the LSTM layer is used to estimate the risk of death in patients in the held-out evaluation dataset by targeting a specified time period (e.g., 30-day or 90-day) when given one clinical note. We calculated positive predictive value as the standard for judging whether obtaining irregular time lapse segments from the LSTM unit is correct or not. Prediction accuracy is calculated as means of comparison between the softmax output (which returns a date range corresponding to the predicted patient death date based on one sample) and a patient's actual death date.

2) A comparison to support vector machine

Partitioning the time dimension is a linear segmentation problem. We use a support vector machines (SVM) as the baseline to calculate prediction accuracy. We compared LSTM with SVM in the classic 70:30 ratio between the training set and validation set. We considered different settings for the initial size of the time segments hyperparameter in our proposed model of 30 days, 90 days, and 360 days.

3) A comparison to the regional classifier

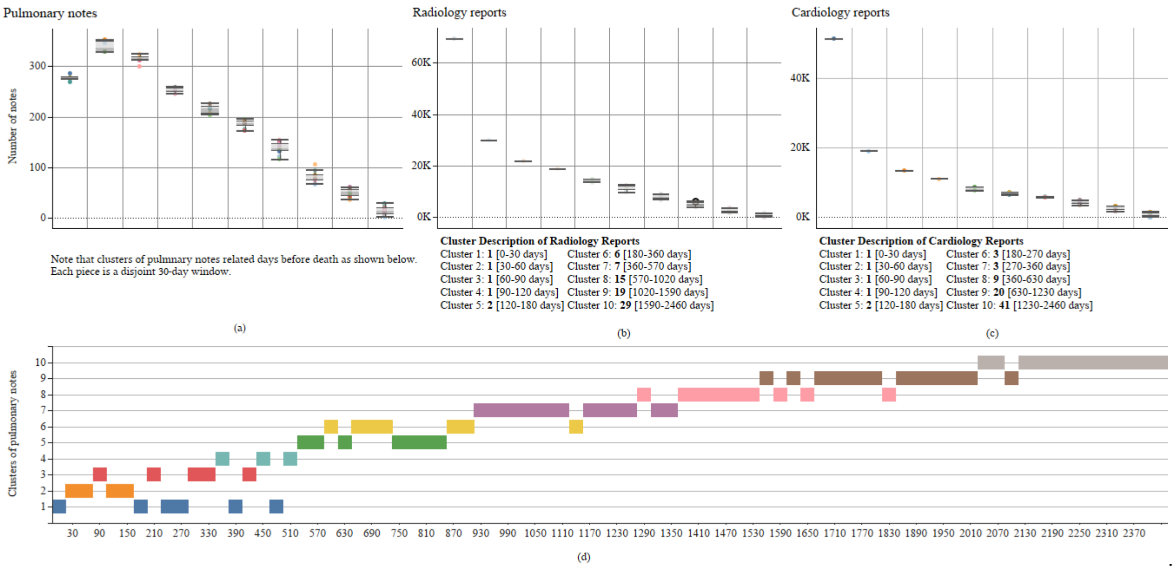
We previously developed a regional classifier based on a spiral timeline for visualizing literature data, which presents research topic words under different themes in a spiral map to show the chronological development of focused research topics [15]. A spiral timeline is able to 1) compactly display the longest possible length of time in a limited space and 2) avoid having a situation where a correlation between two parallel events is missed if all their comparable parameters are similar to each other. When timelines are combined with a geographical map, they depict temporal patterns of events with respect to their spatial attributes [16]. Although the regional classifier cannot be effectively employed because it only considers segments of equal-time (e.g., year), we use it as a baseline.

The regional classifier uses a classic topic model – latent Dirichlet allocation (LDA) [17] to present theme words under different themes. To enhance our themes for COPD, we utilized a representative sentence instead of theme words. More specifically, a representative sentence can be generated by comparing whether the sentence has 3-4 theme words (e.g., 30% of an average sentence length if the entire sentence has 10-14 words) that belong to a specific topic identified by LDA. We used the regional classifier as a baseline to determine the impact of irregular time segments for this task. The goal is to compare the top k representative sentences captured by the regional classifier to our proposed model to determine this impact. We utilize all pulmonary notes from the corpus for this comparison.

IV. Results

A. Raw Data Analysis

Fig. 4 (interactive link for details) shows note distribution when we clustered each type of notes into 10 clusters. Fig. 4 (a), (b), and (c) demonstrate that the number of different types of notes varies in different clusters across different time windows. For example, as shown in Fig 4(c), for cardiology reports, while Cluster 1 only included notes in the 0-30 day time window, Cluster 6 included notes across the 180-270 (i.e., three 30-day) time window. While in general the number of notes in each cluster increased with disease severity toward the death date, a cluster can span multiple inconsecutive time windows (see Fig 4(c)). To overcome this limitation, we therefore developed the a flatten layer and a dense layer (see section E in methods) to handle irregular time lapse segments.



Note that clusters of pulmnary notes related days before death as shown below.
Each piece is a disjoint 30-day window.

(a)

Cluster Description of Radiology Reports

Cluster 1: **1** [0-30 days]
Cluster 2: **1** [30-60 days]
Cluster 3: **1** [60-90 days]
Cluster 4: **1** [90-120 days]
Cluster 5: **2** [120-180 days]
Cluster 6: **6** [180-360 days]
Cluster 7: **7** [360-570 days]
Cluster 8: **15** [570-1020 days]
Cluster 9: **19** [1020-1590 days]
Cluster 10: **29** [1590-2460 days]

(b)

Cluster Description of Cardiology Reports

Cluster 1: **1** [0-30 days]
Cluster 2: **1** [30-60 days]
Cluster 3: **1** [60-90 days]
Cluster 4: **1** [90-120 days]
Cluster 5: **2** [120-180 days]
Cluster 6: **3** [180-270 days]
Cluster 7: **3** [270-360 days]
Cluster 8: **9** [360-630 days]
Cluster 9: **20** [630-1230 days]
Cluster 10: **41** [1230-2460 days]

(c)

(d)

Fig. 4. Distribution of the three types of clinical notes in 10 clusters across different disjoint 30-day windows. Details can be found in the interactive link.

B. LSTM prediction accuracy at mutiple epochs on Dataset 1

Our proposed model outperformed the SVM; for example, it achieved a prediction accuracy of 80.66% on our corpus when setting 90 days as the initial size of temporal segment, compared to the SVM baseline of 7.45% (Table 2).

TABLE II. LSTM PREDICTION ACCURACY AT DIFFERENT EPOCHS COMPARED TO THE BASELINE OF SVMS

Prediction Accuracy (%)	**Initial Size of Time Segment**		
	30-day	***90-day***	***360-day***
1 epoch	0.97	7.74	31.29
10 epochs	29.18	60.13	78.81
23 epochs	71.19	77.92	**78.89**
50 epochs	**78.85**	**80.66**	—
SVM as the baseline	0.83	7.45	27.20

Fig. 5 indicates that the initial size of temporal segment is inversely proportional to the number of training epochs. With the window hyperparameter set to 360 days, our model only performed 23 epochs.

C. A comparison of the spiral timeline

Based on the 50 epochs, we obtained a sequence of time lapse segments from the corpus of pulmonary notes using 90 days as the initial size for each time segment. As shown in Fig. 6, we illustrated the most recent ten time lapse segments prior to death date.

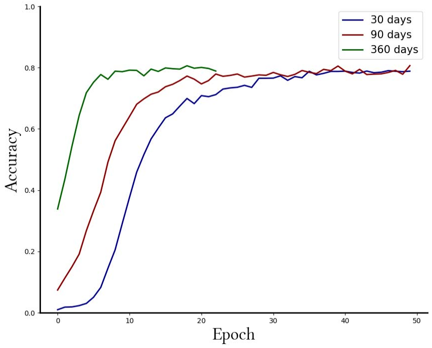


Fig. 5. LSTM Prediction accuracy along a sufficient number of epochs.

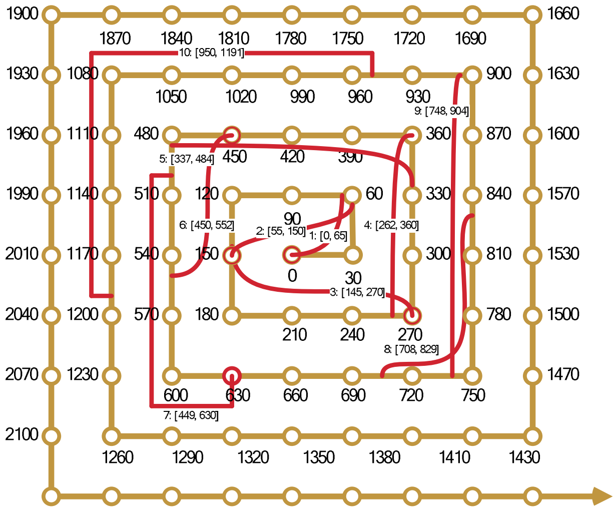


Fig. 6. Visualization of the Regional Classifiers standard spiral timeline (i.e., yellow line with an initial 30-day time window) compared to the first ten irregular time lapse segments (i.e., red line) from our proposed model.

D. *An example comparison of representative sentences on COPD conditions*

We inputted two time lapse segments (i.e., 0 to 65 days before death, 0 to 90 days before death) from Dataset 2 into our baseline LDA model. Note that 0 to 65 days before death is the first time lapse segment which corresponds to the irregular time interval suggested by our proposed model shown in Fig. 5.

Using the LDA model plus the sentence determinate rule, we retrieved the top three generated sentences for COPD conditions by each algorithm (Table 3).

TABLE III. AN EXAMPLE ON REPRESENTATIVE SENTENCES ON COPD CONDITIONS GENERATED BY PROPOSED MODEL COMPARTED TO THE BASELINE OF THE REGIONAL CLASSIFIER

0 to 65 days before death (proposed model)	**0 to 90 days before death (regional classifier)**
These data demonstrate a very severe (FEV1<35% of predicted) obstructive ventilatory deficit. Arterial blood gases reveal respiratory alkalosis with a widened A-a gradient for oxygen. These data suggest a very severe (FVC between 35 and 49% of predicted) restrictive ventilatory deficit.	These data demonstrate a very severe (FEV1<35% of predicted) obstructive ventilatory deficit. These data suggest a severe (FVC between 35 and 49% of predicted) restrictive ventilatory deficit. These data demonstrate a mild restrictive ventilatory deficit.

Note that underlined words were produced by LDA automatically as theme words.

We found that the three sentences generated by our proposed model are all in line with the COPD IV stages of very severe based on the GOLD criteria "FEV1/FVC<0.70" or "FEV1<30% predicted or FEV1<50% predicted plus respiratory failure." In contrast, the three sentences produced from the regional classifier described COPD at different stages, as elicited by the theme words: "very severe," "severe," and "mild."

V. DISCUSSION

Our main finding is that it is feasible to utilize our proposed solution for modeling and predicting COPD progression. Compared to the support vector machine and regional classifier, our empirical study demonstrates that the time segments produced by our proposed model are more interpretable, accurate, and reliable in estimation of COPD mortality. What is surprising is that the accuracy value for the initial size of 90 days was >7 times better than the baseline SVM model after only 10 epochs. This is surprising because it suggests that a the number of execution epochs necessary to outperform baseline models is smaller than we might expect, meaning that the computational load necessary to implement these algorithms into clinical workflows is likely reasonable. Although the 360-day initial time segment didn't finish 50 epochs (it performed 23 epochs), its final accuracy approaches 80% (1.5 times better than 1 epoch). Our results are robust to the initial time segment whatever the size is. Future work can incorporate other risk factors to increase the accuracy of mortality prediction.

The ability to better describe disease progression represents an unmet need for chronic diseases like COPD and would help to inform therapeutic and management choices. The starting point of most existing work aims to find a best-fitting trajectory learned from machine learning models. For example, Wang et al. [4] proposed a probabilistic disease progression model to find the hidden progression process within a learned full progression trajectory. Tilling et al. [10] presented multilevel models to estimate possible therapies from the trajectory of disability progression in relapsing-remitting onset multiple sclerosis. Xiao et al. [5] developed a restricted hidden Markov model to estimate latent states for diabetes using a trajectory learned from the demographic and behavioral data. Although these existing irregular sampling methods exist, they all possess difficulties related to time intervals. There is a lack of methods that can provide substantial interpretation on the mechanism, progression, and key indicators/measurements for the target disease. These general-purpose evidence-based disease progression modeling techniques tend to be specific to the continuous time hypothesis. Thus, they are often used in predicting one patient by fitting a continuous trajectory. In this study, a prediction was generated at the note level (i.e., estimating the risk of death within a time window based on a clinical note) instead of the patient level. The note level is a more actionable level of granularity compared to the patient level.

Chronic diseases usually progress slowly over a long period of time, causing an increasing burden on the patients, their families, and the healthcare system. Our approach may be generalizable to handling longitudinal data (particularly with irregular temporal segments) of other chronic diseases such as Alzheimer's disease and diabetes as a better understanding of their progression may be instrumental for early diagnosis and personalized care.

Another line of efforts, to which our model belongs, is the capture of underlying time dependencies in disease progression using RNN's. The challenge is that time irregularity is common in healthcare applications, but current RNN architectures have challenges in handling it. For example, a temporal pattern with frequent visits or long hospitalizations might indicate a severe progression of the condition whereas a less frequent visit temporal pattern may indicate a stable patient state. A similar idea to deal with irregular time intervals was proposed in [5, 18], with both focused on patient-level data. Different datasets and purposes could lead to differences in study design. The main aim of [5, 18] was to answer the question: how to adjust the existing data to conform to a regular time interval. Pham et al. [18] solved the time irregularities issue between consecutive admissions by setting the forget gate in LSTM to ignore. Baytas et al. [5] addressed the time irregularities issue by modifying the memory cell of LSTM according to the elapsed time.

One limitation of our study is that when merging the three clinical note types, our study only uses a heuristic merger (i.e., we do not consider the relationship among three types of clinical notes to COPD). The three clinical note types might have differences on data correlation for the different stages of COPD disease progression. Learning methods might provide a scoring approach to balance differences (e.g., priority, dataset size) among the three types of clinical notes.

VI. Conclusions

We developed a novel four-layer deep learning model using RNNs to capture irregular time interval. Of which, the output of the LSTM layer aims to estimate all-cause 30-day to 360-day mortality in patients with COPD as the target in this study. Our main finding is that it is feasible to utilize our proposed model for predicting COPD progression, and that this model outperformed baseline models.

References

[1] American Lung Association, "Trends in COPD (Chronic Bronchitis and Emphysema): Morbidity and mortality," March 2013. Avariable from: http://www.lung.org/assets/documents/research/copd-trend-report.pdf [Accessed July 2018].

[2] Global Initiative for Chronic Obstructive Lung Disease, Inc, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease," January 2017. Avariable from: https://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd [Accessed July 2018].

[3] R. G. Barr, D. A. Bluemke, F. S. Ahmed, et al, "Percent emphysema, airflow obstruction, and impaired left ventricular filling," New England Journal of Medicine, 2010, vol. 362(3): pp. 217-227.

[4] X. Wang, D. Sontag, F. Wang, "Unsupervised learning of disease progression models," KDD'14 Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, USA, pp. 85-94, August 2014.

[5] H. Xiao, J. Gao, L. Vu, et al, "Learning temporal tisheng diabetes patients via combining behavioral and demographic data," Kdd'17 Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Canada, pp. 2081-2089, August 2017.

[6] I. M. Baytas, C. Xiao, X. Zhang, et al, "Patient subtyping via time-aware lstm networks," Kdd'17 Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Canada, pp. 65-74, August 2017.

[7] R. A. Pauwels, A. S. Buist, P. M. A. Calverley, et al, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary," American journal of respiratory and critical care medicine, 2001, vol. 163(5): pp. 1256-1276.

[8] C. Che, C. Xiao, J. Liang, et al, "An RNN Architecture with Dynamic Temporal Matching for Personalized Predictions of Parkinson's Disease," SIAM International Conference on Data Mining (SDM 2017), SIAM, USA, pp.198-206, April 2017.

[9] Z. Che, S. Purushotham, K. Cho, et al, "Recurrent neural networks for multivariate time series with missing values," Sientific Reports, vol. 8(1), pp. 6085, April 2018.

[10] K. Tilling, M. Lawton, N. Robertson, et al, "Modelling disease progression in relapsing-remitting onset multiple sclerosis using multilevel models applied to longitudinal data from two natural history cohorts and one treated cohort," Health Technology Assessment (Winchester, England), vol. 20(81), pp. 1-48, October 2016.

[11] B. R. Celli, C. G. Cote, J. M. Marin, et al, "The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease," New England Journal of Medicine, 2004, vol. 350(10): pp. 1005-1012.

[12] T. Mikolov, K. Chen, G. Corrado, et al, "Efficient estimation of word representations in vector space," arXiv preprint arXiv: 1301.3781, 2013.

[13] Y. Bengio, P. Simard, P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," IEEE transactions on neural networks, vol. 5(2), pp. 157-166, March 1994.

[14] S. Hochreiter, J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9(8), pp. 1735-1780, 1997.

[15] C. Tang, H. Zhang, K. H. Lai, et al, "Developing a regional classifier to track patient needs in medical literature using spiral timelines on a geographical map," IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2017). IEEE, USA, pp. 874-879, November 2017.

[16] K. P. Hewagamag, M. Hirakawa, T. Ichikawa, "Interactive visualization of spatiotemporal patterns using spirals on a geographical map," IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC 1999). IEEE, Japan, pp. 296-303, September 1999.

[17] D. M. Blei, J. D. Lafferty, "Dynamic topic models." The 23rd International Conference on Machine Learning (ICML 2006). ACM, USA, pp. 113–120, June, 2006.

[18] T. Pham, T. Tran, D. Phung, et al, "Deepcare: A deep dynamic memory model for predictive medicine," The 20th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD 2016). New Zealand, pp. 30-41, April 2016.