

Original Paper

Asthma Exacerbation Prediction and Risk Factor Analysis Based on a Time-Sensitive, Attentive Neural Network: Retrospective Cohort Study

Yang Xiang¹, PhD; Hangyu Ji², MD; Yujia Zhou¹, MBBS; Fang Li¹, PhD; Jingcheng Du¹, PhD; Laila Rasmy¹, MSc; Stephen Wu¹, PhD; W Jim Zheng¹, PhD; Hua Xu¹, PhD; Degui Zhi¹, PhD; Yaoyun Zhang¹, PhD; Cui Tao¹, PhD

¹School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States

²Division of Gastroenterology, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China

Corresponding Author:

Cui Tao, PhD

School of Biomedical Informatics

The University of Texas Health Science Center at Houston

UCT 600, 7000 Fannin Street

Houston, TX, 77030

United States

Phone: 1 713 500 3981

Email: Cui.Tao@uth.tmc.edu

Abstract

Background: Asthma exacerbation is an acute or subacute episode of progressive worsening of asthma symptoms and can have a significant impact on patients' quality of life. However, efficient methods that can help *identify personalized risk factors and make early predictions* are lacking.

Objective: This study aims to use advanced deep learning models to better predict the risk of asthma exacerbations and to explore potential risk factors involved in progressive asthma.

Methods: We proposed a novel time-sensitive, attentive neural network to predict asthma exacerbation using clinical variables from large electronic health records. The clinical variables were collected from the Cerner Health Facts database between 1992 and 2015, including 31,433 adult patients with asthma. Interpretations on both patient and cohort levels were investigated based on the model parameters.

Results: The proposed model obtained an area under the curve value of 0.7003 through a five-fold cross-validation, which outperformed the baseline methods. The results also demonstrated that the addition of elapsed time embeddings considerably improved the prediction performance. Further analysis observed diverse distributions of contributing factors across patients as well as some possible cohort-level risk factors, which could be found supporting evidence from peer-reviewed literature such as respiratory diseases and esophageal reflux.

Conclusions: The proposed neural network model performed better than previous methods for the prediction of asthma exacerbation. We believe that personalized risk scores and analyses of contributing factors can help clinicians better assess the individual's level of disease progression and afford the opportunity to adjust treatment, prevent exacerbation, and improve outcomes.

(*J Med Internet Res* 2020;22(7):e16981) doi: [10.2196/16981](https://doi.org/10.2196/16981)

KEYWORDS

asthma; deep learning; electronic health records; health risk appraisal

Introduction

Background

Asthma is a common and serious health problem that affects 235 million people worldwide [1] and an estimated 26.5 million people (8.3% of the US population) in the United States [2]. Asthma takes a significant toll on the population, which imposes an unacceptable burden on health care systems. In 2013, the total annual cost of asthma was US \$81.9 billion in the United States [3]. If not well controlled or stimulated by specific risk factors, asthma may develop into exacerbations (asthma attacks), which are acute or subacute episodes characterized by a progressive increase in one or more typical symptoms of asthma (dyspnea, coughing, wheezing, and chest tightness) [4]. In 2016, 12.4 million current asthmatics (46.9%) in the United States had at least one asthma exacerbation in the previous year [2]. Exacerbations of asthma can be severe and require immediate medical interventions, either as an emergency department (ED) visit or admission to the hospital [5]. Serious asthma exacerbations may even result in death [6]. Therefore, it is of practical significance to make early predictions such that interventions can be carried out in advance to reduce the probability of an exacerbation.

Investigations on risk factor analysis or prediction for asthma exacerbation have been respectable, in which the mainstream adopts traditional statistical methods, such as logistic regression [7-10], proportional hazards regression [11], and generalized linear mixed models [12]. However, most of them have only explored a small group of candidate risk factors and are usually hard to extend to other data sets and make personalized predictions difficult [13,14]. With the explosion of health care data in recent years, machine learning methods have taken a nontrivial place in this domain, benefiting from their general superiority over statistical methods in processing larger numbers of variables and flexibility in modeling more complex correlations [15]. Typical models include naïve Bayes [16], Bayesian networks [16-19], artificial neural networks [17], Gaussian process [17], and support vector machines [16,17]. However, although different attempts have been made, there are still several deficiencies in traditional machine learning methods. For example, ignoring temporal dependencies between variables might not provide a meaningful risk estimation of future exacerbations for individual patients [14]. Furthermore, most approaches only concentrate on the quantitative evaluation of prediction performance, but lack further attention to personalized risk factors [20].

Recent revolutions in health artificial intelligence started from deep learning, which has an upper hand on health care predictions because of its flexibility in dealing with longitudinal data [21], powerful learning capabilities [22], and ability to alleviate the problem of data irregularity [23]. One of the most popular architectures is recurrent neural networks (RNNs), which make predictions according to the sequence of historical events. Dozens of successes have been achieved in applying deep learning to disease predictions [24], mostly using variants of RNNs with distinct network components, for example, by adding an attention mechanism to evaluate the weights of each

variable [25-29] or by using special configurations to tackle the problem of time decays [23,25,27,30-32]. Typical prediction tasks include the prediction of diabetes mellitus [23], Parkinson disease [29,33], chronic heart failure [26], sepsis [34], mortality, and readmission [25]. However, deep learning-based studies on the prediction of asthma exacerbation remain lacking. Do et al [35] proposed a protocol for the prediction based on RNNs and reinforcement learning but did not test the method on real-world data.

Objectives

Inspired by previous studies, we applied long short-term memory (LSTM) [33], a popular RNN variant commonly used by previous predictive models [23-25,29,34] as the main framework for asthma exacerbation prediction, which can mitigate the gradient vanishing problem in RNNs. We proposed the time-sensitive, attentive neural network (TSANN), which employs a self-attention mechanism [36] to help model the context of both visit-level and code-level variables. Meanwhile, to incorporate the impact of elapsed time, we projected the relative time of each clinical variable into a low-dimensional space and combined it with the code representations. Using the attention weights of the TSANN, data analysis was then conducted to investigate personalized and cohort-level risk factors.

There are major differences between TSANN and recent state-of-the-art deep learning-based clinical predictive models such as time-aware LSTM (TLSTM) [23], Reverse Time Attention model (RETAIN) [27], and Attention-based Time-aware Disease Progression (ATTAIN) [32]. First, the model structures are different. Compared with TLSTM and ATTAIN, which only include 1 layer of RNN, our two-layer architecture enables us to analyze the relative importance of each event within each visit. Although RETAIN also has 2 layers of attention, it does not have explicit hierarchical structures as TSANN. Instead, an additional inference step is required to obtain the contribution of each variable. Second, TSANN uses a different approach to model the elapsed time. RETAIN, TLSTM, and ATTAIN feed the time elapsed into a decay function as a single value and multiply it with the network memory. In comparison, the elapsed time embeddings in TSANN are more analogous to position embeddings in natural language processing, which were introduced to model the relative distance between words by learning multidimensional and semantic representations to facilitate certain tasks such as relation classification [37] and neural language modeling [36,38,39]. By using time embeddings, we assume that time is no longer a single value as it was used in previous methods, but it can represent more complex patterns together with clinical variables such as varying lengths of correlations between variables.

The primary aims of this study were (1) to propose a novel predictive model with better performance and (2) to add the transparency of the model by visualizing contributing factors at both the individual and cohort levels. Furthermore, the proposed model can potentially be applied to other clinical problems. Deep learning models are usually scalable. Although focusing on asthma exacerbation for this specific project, the

proposed approach can also be adopted in risk predictions for other chronic diseases. We hope that the associated pipeline of deep learning-based predictive modeling, including data collection, model training, model evaluation, and risk factor analysis, can help the clinical community better understand the underlying mechanisms of disease progression and assist in decision making.

Methods

Problem Statement

Given a sequence of historical clinical variables in patients with asthma, we aimed to evaluate the risk of developing asthma exacerbation in the designated time window. Meanwhile, personalized contributing factors are to be identified to facilitate the evaluation of disease progression and make early interventions.

Database

The study used Cerner Health Facts, a Health Insurance Portability and Accountability Act-compliant database collected from multiple enrolled clinical facilities, containing mostly inpatient data. Data in Health Facts were extracted directly from the electronic health records (EHRs) from hospitals with which Cerner has a data use agreement. Encounters may include the pharmacy, clinical and microbiology laboratory, admission, and billing information from affiliated patient care locations. All personal identifying information of the patients was anonymized. In this study, we primarily focused on the impact of clinical factors on asthma exacerbation; therefore, we extracted diagnoses, medications, and demographic characteristics such as gender, race, and age from the database as clinical variables or clinical events. The University of Texas Health Science Center (UTHealth) had agreements with Cerner to use these data for research purposes. The institutional review board at the UTHealth approved the study protocol.

Study Design

We conducted a retrospective study to predict the risk of asthma exacerbation. Patients' records between 1992 and 2015 were extracted from the Cerner database. For clarity, we defined several terms in advance (Table 1).

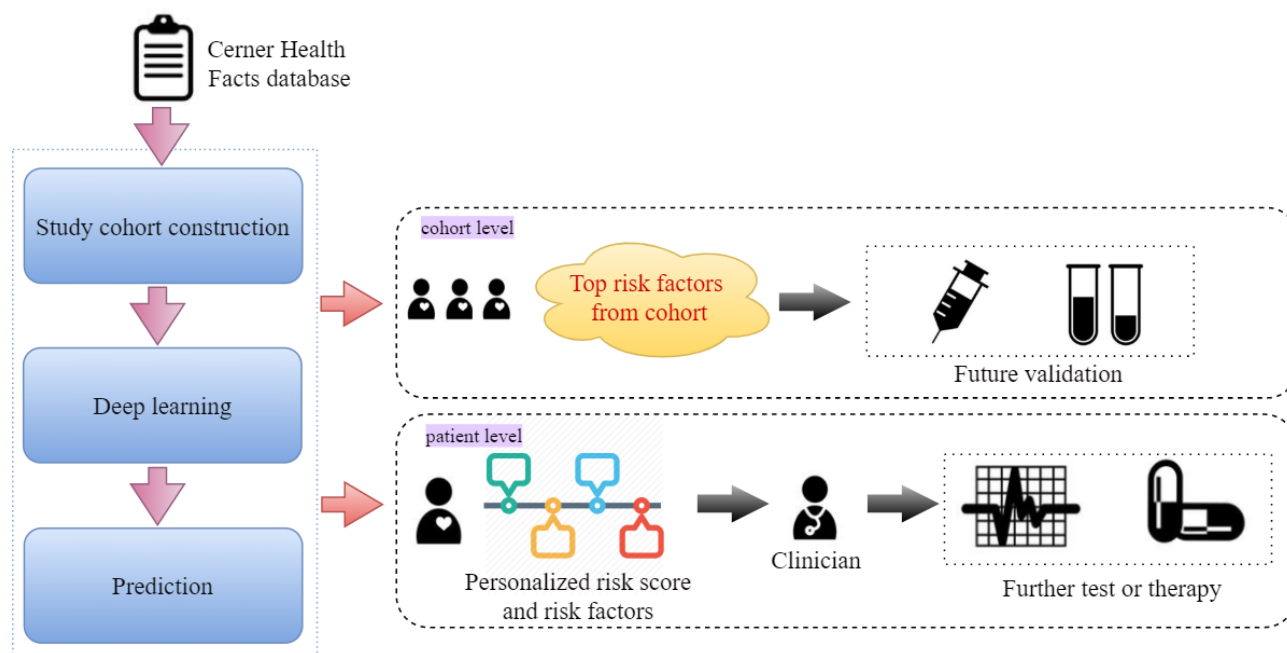
We built 2 cohorts as simulations for both the real-world application scenario (*early prediction*) and the model evaluation (called *next-visit prediction* in many previous studies [27,32]). In early prediction, we could not foresee when the exacerbation would happen but could only evaluate the future risk at each visit. In our study, we selected the fifth visit from the asthma index as the prediction date (*testing set A*) according to the average number of visits (5.78, SD 6.04) between asthma index and exacerbation among the patients. The detailed steps for the cohort selection are listed in [Multimedia Appendix 1](#) [4,25,40-55]. In next-visit prediction, we simply set the penultimate visit as the prediction date (*testing set B*). We set testing set A as our primary evaluation set as it was much closer to the realistic diagnostic situation.

The TSANN model was trained to evaluate the risk of asthma exacerbation given the observed time window. The main outcomes of the proposed method are (1) a score that measures the risk of asthma exacerbation for each patient and (2) visualization of the results, including a personalized heatmap identifying the importance of each clinical variable in the observed time window, cohort-level risk factors, and their temporal distributions among patients. On the basis of the outcomes, further data mining or clinical trials can be carried out for validation. For example, cohort-level factors will help data scientists reduce labor and expertise in collecting candidate risk factors from the literature before conducting a regression analysis. Patient-level factors will facilitate physicians and patients in better understanding disease progression. The workflow of this research is shown in [Figure 1](#).

Table 1. Defined terms for asthma exacerbation prediction.

Term	Definition
Index date	The date of the first diagnosis of asthma in a patient's EHR ^a
Exacerbation date	The date of the first diagnosis of asthma exacerbation after the index date
Case group	Patients with asthma and later asthma exacerbations within 365 days and satisfying the inclusion and exclusion criteria
Control group	Patients with asthma but without exacerbations within 365 days and satisfying the inclusion and exclusion criteria
Prediction date	Training set: for the case group, the visit date before the exacerbation date; for the control group, the penultimate visit date within 365 days: <ul style="list-style-type: none"> • Testing set A: the fifth visit starting from the index date • Testing set B: defined analogously to the training set
Observed time window	The time window between the index date and the prediction date

^aEHR: electronic health record.

Figure 1. The workflow of the prediction and risk analysis of asthma exacerbation.

Selection of Study Subjects

The subjects in the study were patients with a diagnosis of asthma. The inclusion and exclusion criteria derived from previous studies [4,56] were as follows.

Inclusion Criteria

The subjects in the study were patients with at least one record of asthma diagnosis. The definitions of asthma and exacerbation are as follows.

Asthma

1. Asthma diagnosis codes were provided according to the International Classification of Disease Code (ICD; ICD-9 code 493.xx or ICD-10 code J45.xx). This is the first occurrence of asthma in the patient's EHR.
2. At least one of the asthma medications was prescribed on the asthma diagnosis date (the index date). Asthma medications include short-acting beta agonists, inhaled corticosteroids (ICS), long-acting beta agonists (LABA), leukotriene receptor antagonists, anticholinergics, and ICS/LABA combinations.

Asthma Exacerbation

1. Asthma (ICD-9 code 493.xx or ICD-10 code J45.xx) was given as a primary diagnosis for an ED visit or hospitalization.
2. At least one oral corticosteroid treatment was received.

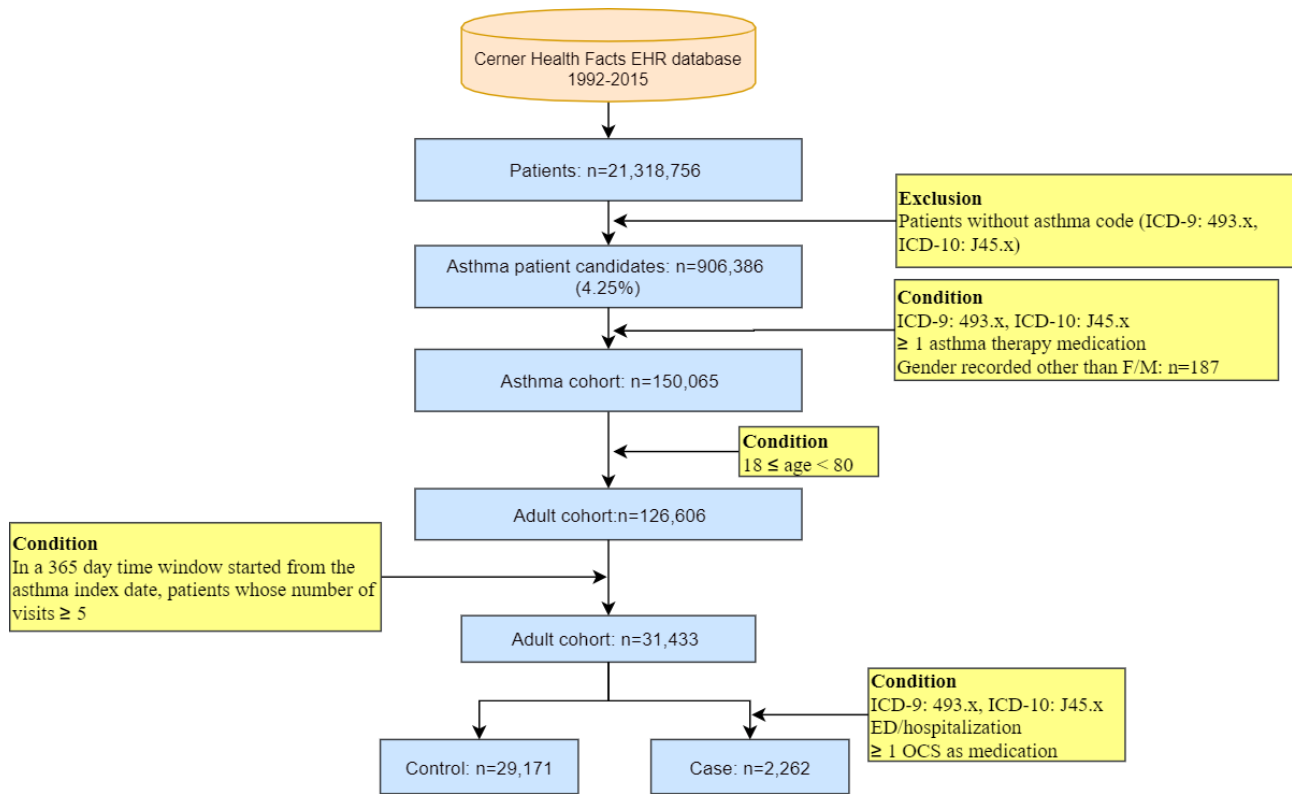
Exclusion Criteria

To allow the data to better fit for machine learning models, we excluded the following patients:

1. Those with missing or unclear time information (eg, with a wrongly recorded format of time stamps)
2. Those with a gender other than male or female
3. Those whose number of visits is <5 in the observed time window

This study only focused on adult patients aged between 18 and 80 years. In the end, 31,433 individuals remained, including 2262 cases and 29,171 controls (case by control ratio approximately 1:13). The cohort selection process is shown in [Figure 2](#). A detailed descriptive analysis of the cohort is shown in [Multimedia Appendix 1](#).

Figure 2. Cohort selection process for the study of asthma exacerbation. ED: emergency department; EHR: electronic health record; ICD: International Classification of Disease Code; OCS: oral corticoids.



Time-Sensitive Attention Neural Network

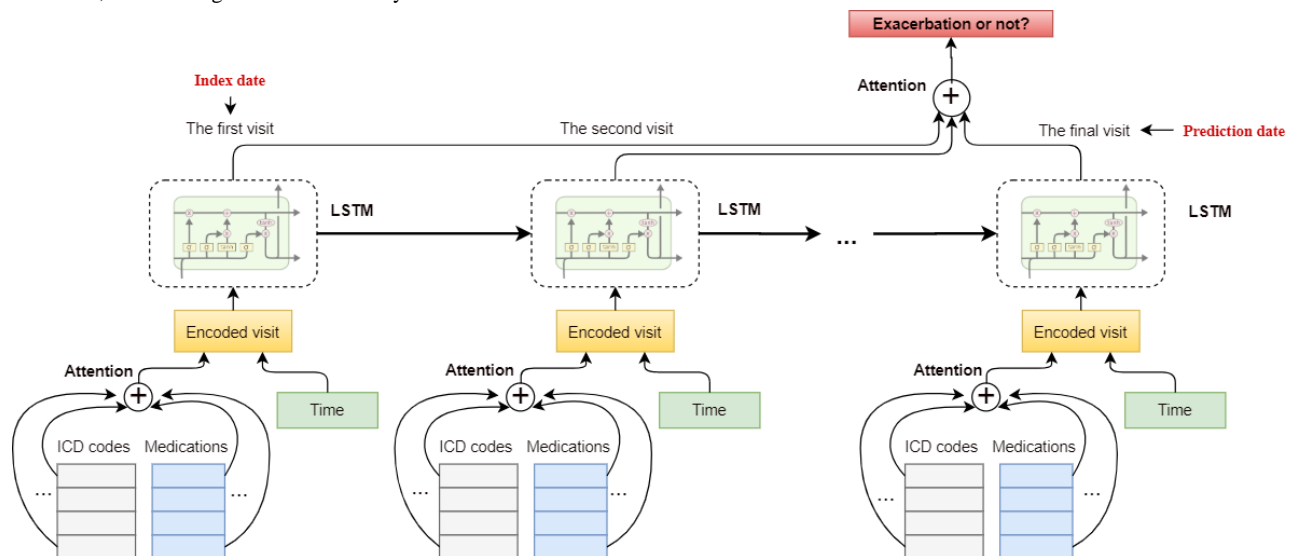
Model Overview

TSANN takes the whole sequence of clinical variables in the observed time window as inputs and outputs the probability of asthma exacerbation (Figure 3). The architecture of TSANN is based on LSTM and strengthened by the addition of hierarchical attention and elapsed time embeddings.

For each visit, multiple clinical variables were encoded in the input layer and averaged through the code-level attention

mechanism. The elapsed time embedding is attached to each visit as complementary information to indicate the time interval between the date of each visit and the prediction date. LSTM then accepts the sequence of encoded visits as inputs and outputs further encodings for each visit. The visit-level attention layer is then applied to the outputs of the LSTM to summarize all the visits for each patient. Finally, by feeding the output of visit-level attention into the Softmax function, a probability indicating the risk of disease onset is generated.

Figure 3. Overview of the time-sensitive attentive neural network model for asthma exacerbation prediction. ICD: International Classification of Disease Code; LSTM: long short-term memory.



Input

The inputs of the model consist of 2 types of features. One type is clinical concepts (we use *clinical concepts* and *clinical variables* interchangeably), including ICD codes, medications, and demographic features. All ICD-10 codes were converted into ICD-9 based on predefined mappings [57] because very few diagnosis codes in our data set were encoded by ICD-10 as the data collection time range is between 1992 and 2015, but the implementation of ICD-10 started in October 2015. All medications were normalized to their generic names. The demographic features included age, gender, and race, which were only taken as inputs on the prediction date. Using a projection matrix $W_{cemb} \in \mathbb{R}^{V_c \times D_c}$ (V_c : concept vocabulary size and D_c : concept embedding dimension), we mapped each clinical concept into a concept-embedding vector:

$$C_{ij} = W_{cemb} \cdot v_{ij}$$

where C_{ij} is the generated concept-embedding vector and $v_{ij} \in \mathbb{R}^{D_c}$ is the one hot vector denoting the existence of clinical concept j in visit i .

The other feature type is time features, which indicates the occurrence time for each clinical variable. Intuitively, variables with different time stamps would behave differently in prediction. For instance, in many cases, a clinical event that happened several days ago would play a more important role than one that happened several months ago. Meanwhile, due to the nature of data irregularity and deficiency of EHRs, successive visits always have diverse time intervals [23], which makes it indispensable to consider the time elapsed when conducting predictive modeling.

Elapsed time embeddings were introduced to represent the relative time gap for each clinical concept. Specifically, taking the time of the prediction date T_0 as a pivot, the time attribute of each clinical concept is the absolute difference between its occurrence time T_i and T_0 , that is, the relative time gap $T_0 - T_i$. As the observed time window has an upper bound of 365 days, the vocabulary size V_t of the time embeddings was set as 365.

We applied a matrix $W_{temb} \in \mathbb{R}^{V_t \times D_t}$ to project each time value to an m -dimension vector. Unlike the clinical concept embeddings, elapsed time embeddings are fed into the model after the code-level attention and assigned to each visit. The equation to obtain the elapsed time embedding for each visit is analogous to that for concept embeddings, where: $t_{ij} \in \mathbb{R}^{D_t}$

$$T_i = W_{temb} \cdot t_i$$

Specifically, the minimum time unit in this study was set as *day*.

Code-Level Attention

Attention is a mechanism specifically designed for deep neural networks that acts as an information filter; meanwhile, it can alleviate information loss when dealing with long sequences. It selects important sequence steps by assigning them different weights [58,59]. Through attention, each clinical concept is

assigned a weight such that important concepts would have larger weights than the others. We adopted the attention mechanism from Yang et al [60], in which the weight of each variable is generated according to the sequence and context vector. Specifically, given the set of codes $v_{ij}, j \in \{1, 2, \dots, M\}$ in the i th visit, the encoded representation for visit v_i can be generated by:

$$u_{ij} = \tanh(W_v v_{ij} + b_v)$$

$$\beta_{ij} = \frac{\exp(u_{ij}^T u_v)}{\sum_k \exp(u_{ik}^T u_v)}$$

$$v_t = \sum_j \beta_{ij} \cdot v_{ij}$$

where W_v and b_v are the weight and bias for matrix transformation, u_{ij} is the attention vector for each code j in v_i , u_v is the context vector for v_i , which is randomly initialized and updated during training, and β_{ij} is the attention weight for the concept V_{ij} based on which we can generate its final weight. By assigning time embeddings to the i th visit v_i , the representation of each visit is updated as $v_i \leftarrow v_i \oplus T_i$, where \oplus denotes the matrix concatenation.

Visit-Level Attentive LSTM Layer

Taking the encoded representation of each visit as input, LSTM models the sequential information in the observed time window and obtains the summarization at the final step (the prediction date). The advantage of LSTMs over traditional RNNs is that they can alleviate the gradient vanishing problem and are thus able to retain longer *memories* from prior time stamps [61,62]. LSTMs are implemented by several matrix multiplications and nonlinear transformations that aim to mimic the memory mechanism of human brains, which are called *gates*, signifying that the network can select effective information and abandon useless information. The equations of the LSTMs are as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, v_t] + b_f)$$



$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, v_t] + b_c)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, v_t] + b_o)$$



where W_s and b_s are the weights and biases for different gates or cells (f_i : forget gate, i_i : input gate, C_i : memory cell, o_i : output

gate, and h_t : hidden cell) and σ is the activation function, such as *Tanh* or *sigmoid*.

By assigning attention weights to the outputs of LSTM from each step, we can weight each visit in the observed time window and obtain a summary of these visits as r_p :

$$u_i = \tanh(W_p v_i + b_p)$$

$$\alpha_i = \frac{\exp(u_i^T u_p)}{\sum_k \exp(u_k^T u_p)}$$

$$r_p = \sum_j \alpha_j \cdot v_j$$

where W_p and b_p are the weight and bias for matrix transformation, u_i is the attention vector for each visit i given v_i , u_p is the context vector, and α_j is the attention weight for each visit v_j . This process can be seen as a simulation of the diagnosis

procedure of a clinic visit, during which a physician would look back into a patient's EHR, measure the impacts of each historical clinical event, and make the final decision.

Output

The visit-level attention layer compresses all the information in the observed time window into a fixed-length vector r_p . The output of attention goes through a fully connected layer with nonlinear activation. Finally, a Softmax function is applied to generate the prediction probability, P :

$$p = \text{softmax}(W_c \cdot r_p + b_c)$$

where P is used as the score to evaluate the risk of developing asthma exacerbation.

Evaluation

Area under the receiver operating curve (AUC) is widely used as an evaluation metric for predictive models, which reflects a balance between sensitivity and specificity [63]. According to the prediction probability P (between 0 and 1) for each instance, the AUC value is generated by setting different cutoffs. The methods listed in Table 2 were compared in our experiments.

Table 2. The methods used for comparisons.

Method	Note
LR ^a	A popular conventional machine learning algorithm [64], usually serving as a strong baseline in predictive modeling [27]. The input of LR is a fixed-length feature vector that denotes the frequencies of each variable. For LR considering time, we associate each variable with its time stamp and expand the vocabulary. We did not use <i>day</i> as the time unit as it would have introduced a greater number of variables (ie, 12,390×365 [the code vocabulary size×the maximum number of days]), which would have been too sparse and difficult for computation. Instead, we set month as the time unit, and finally, 148,680 distinct clinical variables were generated. We employed the Synthetic Minority Oversampling Technique [65] to help alleviate the problem of data imbalance
MLP ^b	The MLP model used in this study contains 1 input layer and 1 Softmax layer [66]. The representations of all the codes were averaged on each dimension after being projected to the embedding space for each patient
LSTM ^c	The basic LSTM algorithm, taking the sequence of the clinical variables as input ordered by time. The variables in each visit are averaged
ALSTM ^d	Comprising 1 layer of LSTM and 1 layer of attention
TLSTM ^e [23]	The time-aware LSTM model, which is one of the state-of-the-art predictive models. In TLSTM, the time gap is used to compute the information decay in the LSTM unit
RETAIN ^f [27]	A two-layer attention model, which is another state-of-the-art model for the prediction of disease onset. In RETAIN, the time features are not embedded as vectors but real values denoting the gaps from the first visit
ATTAIN ^g [32]	A modification of TLSTM with special types of attention mechanisms added (flexible attention). It also uses a similar time decay function as RETAIN. We implemented it ourselves using TensorFlow
TSANN ^{h-I}	The proposed TSANN model but with the second attention layer removed. Prediction is based on the final state of LSTM
TSANN-I-step	Apply the time-encoding method from Song et al [39] on TSANN-I. In TSANN-I-step, although time was also encoded using a vector, it only showed the order of each visit, for example, 1, 2, 3 for consecutive visits, but not the actual elapsed time
TSANN-II	A complete version of the proposed TSANN model

^aLR: logistic regression.

^bMLP: multilayer perceptron.

^cLSTM: long short-term memory.

^dALSTM: attention long short-term memory.

^eTLSTM: time-aware long short-term memory.

^fRETAIN: Reverse Time Attention model.

^gATTAIN: Attention-based Time-Aware Disease Progression.

^hTSANN: time-sensitive attentive neural network.

For evaluation, we first split the data into a training set and a held-out testing set with a ratio of 8:2. Furthermore, five-fold cross-validation was performed on the training data set for parameter tuning. During cross-validation, a grid search was applied to tune the hyperparameters including learning rate (0.0005, 0.001, 0.005, 0.01), l2 penalty (0.0001, 0.0005, 0.001), batch size (32, 64, 128), activation function for LSTM (ReLU [40] and Leaky_ReLU [41]), whether to add batch normalization [45], and the optimizer selection between RMSprop [42] and Adam [43]. We then averaged the AUCs of each epoch (up to 30 epochs) across five folds to obtain the best training epoch. The optimal hyperparameters were adopted to retrain the model on the entire training set and produce the AUC on the testing set. Finally, the hyperparameters for the model TSANN-I, which has the best AUC value, were as follows: batch size=32, concept embedding dimension=100, time embedding dimension=20, Adam as the optimizer with learning rate=0.001, l2 penalty=0.0001 for all parameters, Leaky_ReLU as the

activation function, and adding batch normalization before Softmax. The codes for RETAIN and TLSTM were reused from the respective studies. All other deep learning models were implemented with TensorFlow [44] and trained on Nvidia Tesla V100, Quadro P6000, and Titan XP GPUs. We shared our code on GitHub to facilitate other researchers [67].

Results

AUC Values

AUC values with (+time) and without the consideration of time (-time) on testing set A (the primary evaluation set) are shown in Table 3. In the table, for TLSTM, we only considered a +time version as it is defined as a time-aware variant of LSTM, and for multilayer perceptron (MLP), LSTM, attention long short-term memory (ALSTM), TSANN-I, TSANN-I-step, and TSANN-II, we used the elapsed time embeddings introduced in this study to include time.

Table 3. Area under the curve (AUC) values by different models (-time: time information was excluded and +time: time information was included).

Method	AUC ^a -time	AUC+time
LR ^b	0.6447	0.6773
MLP ^c	0.6545	0.6753
LSTM ^d	0.6045	0.6567
ALSTM ^e	0.6346	0.6714
TLSTM ^f	—	0.6548
ATTAIN ^g	0.6119	0.6597
RETAIN ^h	0.6455	0.6882
TSANN ⁱ -I	0.6692	<i>0.7003</i> ^j
TSANN-I-step	0.6463	—
TSANN-II	<i>0.6827</i>	0.6855

^aAUC: area under the receiver operating curve.

^bLR: logistic regression.

^cMLP: multilayer perceptron.

^dLSTM: long short-term memory.

^eALSTM: attention long short-term memory.

^fTLSTM: time-aware long short-term memory.

^gATTAIN: Attention-based Time-aware Disease Progression.

^hRETAIN: Reverse Time Attention model.

ⁱTSANN: time-sensitive attentive neural network.

^jThe optimal value for each column is italicized.

When comparing vertically (different rows) and considering time information, we noticed that TSANN-I achieved the optimal AUC value, improving the strongest baseline (RETAIN) by 1.21% (the difference was significant according to the Wilcoxon test with $P=.03$). Among other methods, TSANN-II achieved a performance comparable with that of RETAIN. The conventional machine learning method logistic regression (LR) behaved better than some deep learning methods but was worse than RETAIN, TSANN-I, and TSANN-II. TSANN-I-step, which only used time embeddings to denote the relative position of

each visit, did not produce good results. Although TLSTM and ATTAIN performed well on other tasks, they did not obtain satisfactory results on our data. For results without time, TSANN-I and -II performed much better than others, with a maximum improvement of 2.82%.

When comparing the results horizontally (-/+ time), considerable improvements were observed after adding time information on most methods; for example, TSANN-I obtained

a 3.11% improvement. Surprisingly, TSANN-II, when integrating time embeddings, did not improve considerably.

Better performances by TSANN models and considerable improvements after adding the time information could also be observed on testing set B (we did not list the results here but showed them in [Multimedia Appendix 1](#) as it is not our primary evaluation set). As expected, the general results on testing set B were better than those on testing set A, as a sample conveys more complete information.

Patient-Level Risk Factors

In this study, a heatmap was used for each patient’s EHR for the visualization of highly associated variables or possible risk factors. The heatmap illustrates how each variable behaves in each visit during the progression of asthma. Each grid in the heatmap is colored based on the attention weights derived from the model. The darker an area, the more important the clinical variable and the higher the association it has with exacerbation.

For example, [Figure 4](#) shows a case where the symptoms of *hypoxemia*, *shortness of breath*, and *wheezing* (799.02, 786.05, and 786.07 in ICD 9, respectively) were recognized as highly associated variables. A possible explanation might be that the patient’s status of hypoxemia worsened the condition of asthma following symptoms in the breath, and asthma exacerbation was then diagnosed.

These highly associated variables can either be signs of asthma worsening or be triggers for exacerbation, which requires further confirmation by domain experts. Signs including symptoms and treated medications may convey important clues for disease progression and will help clinicians in making final diagnoses, whereas the triggers behaving as personalized risk factors will potentially benefit early interventions. In addition, each heatmap is associated with a probability score derived from equation 15, indicating the risk of the patient in developing an exacerbation (the top row in [Figure 4](#), *predicting 1* indicates *predicting exacerbation*).

Figure 4. An example of a heatmap with highly associated clinical variables, such as hypoxemia (D_799.02), shortness of breath (D_786.05), and wheezing (D_786.07).



Cohort-Level Risk Factors

We also discovered common highly associated variables at the cohort level using the normalized multiplication of the visit-level and code-level attention weights. We then asked the physician to help distinguish the type of these variables according to their expertise and literature lookup. Some of these variables can be confirmed as risk factors, for example, poor control of respiratory diseases [4] and gastroesophageal reflux disease [44], although others need further validations, such as chest pain, migraine, and use of some medications. The top-ranked variables derived from the model are shown in [Table 4](#). The details of the method and the explanations of these factors are described in [Multimedia Appendix 1](#).

Apart from demonstrating the list of cohort-level factors, using the weights generated by equation 2 and equation 3 in [Multimedia Appendix 1](#), we can also visualize how each clinical variable contributes across time; for example, a variable may

behave distinctly among individuals with different action times or different incidences. [Figures 5](#) and [6](#) present 2 examples in which the time distributions for the clinical variables are displayed through scatters. In these scatters, each circle represents a patient where its size and color depth denote the importance of the corresponding variable. In the figures, the x-axis represents the time gap between the occurrence date of the variable and the prediction date, whereas the y-axis is employed merely for cosmesis. We randomly selected a maximum of 2000 patients to plot this figure.

[Figures 5](#) and [6](#) were derived from an ICD code (530.81) and a medication (fentanyl), respectively. We observed different effective time ranges for these 2 factors, where the first factor tends to distribute more intensively between the previous 250 to 50 days, whereas the second factor focuses more intensively on the previous 100 days. We hope that these visualizations can help determine the temporal distributions of highly associated factors to aid asthma control.

Table 4. Clinical variables with the top-ranked weights (N stands for the clinical variable presented in N months before the prediction date).

Sr. No.	ICD ^a -9/occurrence time	Medication/occurrence time
1	493.9×asthma/0-5 ^b (meaning diagnosed with asthma multiple times before exacerbation)	Methylprednisolone/0, 1 ^c
2	786.07 wheezing/0-2 ^d	Prednisone/0, 1, 2 ^c
3	496.0 chronic airway obstruction not elsewhere classified/0, 1 ^e	Ipratropium/0, 1, 2 ^c
4	530.81 esophageal reflux/0 ^b	Midazolam/0, 1, 2 ^d
5	V46.2 dependence on supplemental oxygen/0 ^d	Hydromorphone/0-2 ^e
6	787.02 nausea alone/0 ^d	Heparin/0, 1 ^d
7	786.50 unspecified chest pain/0 ^d	Acetaminophen-oxycodone/0 ^b
8	V08 HIV infection status/0 ^e	Fentanyl/0 ^e
9	786.59 other chest pain/0 ^d	Methylprednisolone/2-4 ^e
10	786.05 shortness of breath/0 ^d	Glycopyrrolate/0 ^b
11	V58.69 long-term (current) use of other medications/0 ^e	Lidocaine/0 ^d
12	784.0 headache/0 ^e	Dexamethasone/0 ^d
13	346.90 migraine, unspecified, without mention of intractable migraine without mention of status migrainosus/0 ^e	Promethazine/0 ^d
14	V58.66 long-term (current) use of aspirin/0 ^b	Atorvastatin/0 ^d
15	491.21 obstructive chronic bronchitis with (acute) exacerbation/0 ^e	Furosemide/0 ^c

^aICD: International Classification of Disease Code.

^bIdentified possible risk factors of asthma exacerbations by the domain expert. The authors regard these as containing valuable information.

^cThese medications can be used to treat asthma or control the symptoms of asthma. In this study, it was difficult to determine whether these medications are risk factors as we were unable to investigate the dosage of these medications in the current study. Inappropriate medication use, short-acting beta agonists/inhaled corticosteroids, could also lead to asthma exacerbations.

^dThese factors were symptoms, comorbidities, or combined medications. We believe they were not risk factors for asthma exacerbations.

^eIt could hardly be determined whether these factors caused asthma exacerbations, but they demonstrated high associations. The authors regard these as containing valuable information.

Figure 5. Time distribution of the contribution of the clinical variable gastroesophageal reflux disease is denoted by ICD-9:530.81. ICD: International Classification of Disease Code.

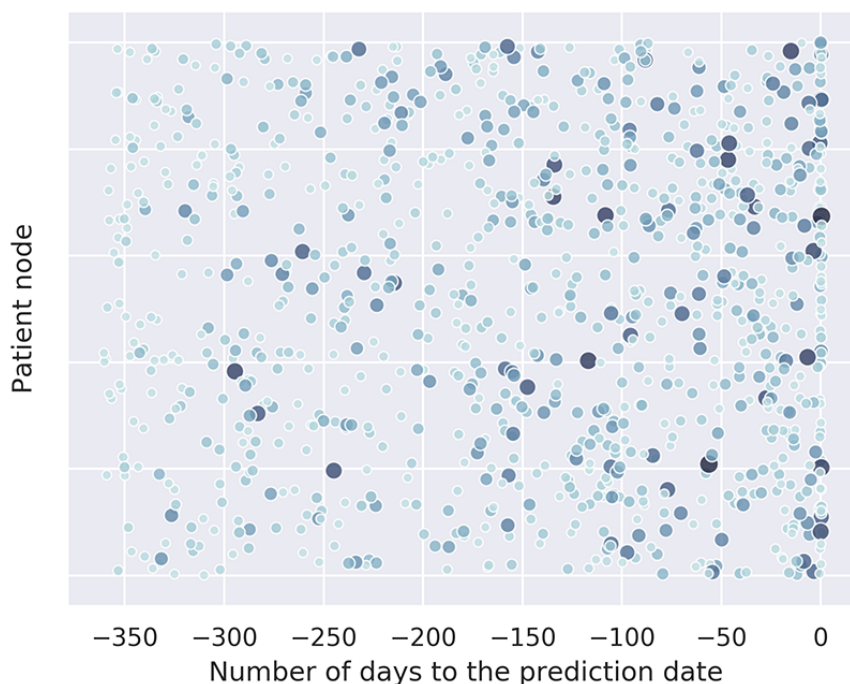
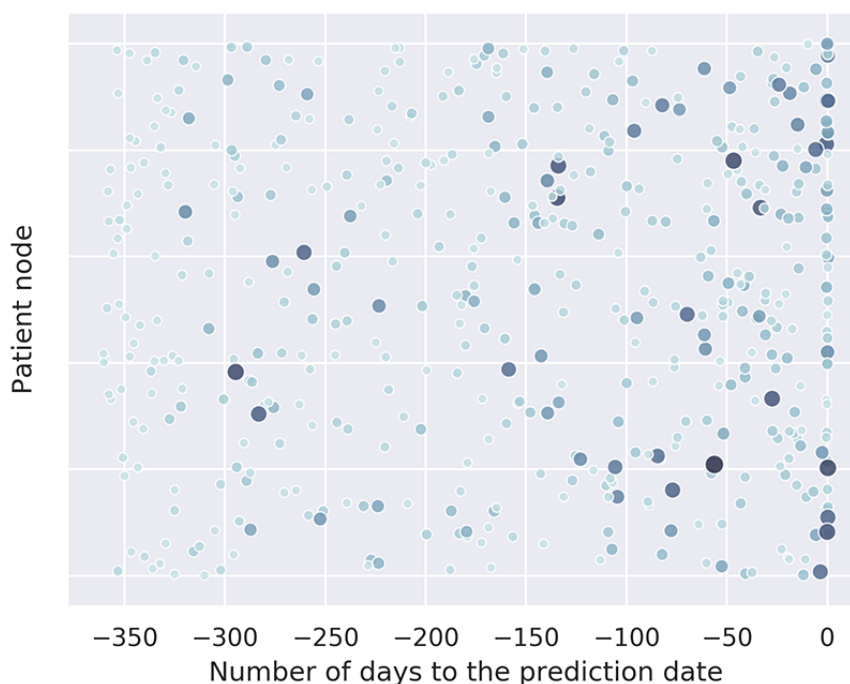


Figure 6. The time distribution of the contribution of the clinical variable fentanyl.



Discussion

Principal Findings

Our proposed method obtains the optimal AUC value on the prediction task, with hierarchical attention and elapsed time embeddings as its booster. The visualization also provides useful tracks for a better understanding of disease progression. The primary outcomes of this study are as follows: (1) a new state-of-the-art predictive model for asthma exacerbation prediction was proposed and validated and (2) a reasonable pipeline of disease risk prediction and factor analysis was

introduced. Some of the identified risk factors can be validated from the literature, which shows the effectiveness of the method, whereas some other factors, although supportive shreds of evidence were seldom reported in previous studies, offer meaningful insights for further research. The discussions in this section are primarily based on the results of testing set A without additional comments.

Model Performance

TSANN-I and TSANN-II have the capacity to capture nonlinearities and learn more complex dependency relationships between variables, benefiting from the structure of hierarchical

attention and the addition of elapsed time embeddings. It is difficult for LR and MLP to learn temporal dependencies between variables according to their natural structures, which makes them perform worse than TSANNs. However, they obtained better AUCs than LSTM and ALSTM on testing set A, partly because the truncation of EHRs by 5 visits weakens their advantages in modeling longer sequences. Although TLSTM and ATTAIN also integrated the time information, they did not obtain satisfactory results as other methods. It is likely that the combination of the single-layer deep learning structure and the numerical time decay function is insufficient in dealing with more complex temporal patterns on this data set and even confuses the classifiers. RETAIN also has a hierarchical attention structure, as introduced in the Introduction section. However, one of its attention mechanisms is applied to each code-embedding dimension, which is different from ours and requires an additional inference step for interpretation. In addition, our addition of time embeddings enhanced the flexibility of modeling the time information, which contributed considerably to the performance.

A typical characteristic of the EHR data is irregularity, which means that the time gaps between clinic visits are irregular and the visits are often sparsely distributed along the timeline and sometimes are even missing. Thus, the predictive model is responsible for serializing the visits for each patient with consideration of time elapses between continuous visits and reduces the effect of missing data. The comparisons between results with and without time information in Table 3 demonstrate the effectiveness of considering time elapses in this cohort. It might be inferred that the prediction of asthma exacerbation is quite time sensitive and most of the critical risk factors should have been time stamped. For instance, even for a visit just before the prediction date, if its occurrence is several months earlier, its impact would be reduced. Similar cases can also be found in Baytas et al [23], who reported an improvement of 6% from LSTM to TLSTM. In comparison, for TSANN-I-step, although time embeddings were also used, they were only used to denote the relative position of each visit in the sequence but lacked the ability to represent time decays, which can hardly obtain satisfying results here. However, adding time to TSANN-II did not improve much as in other methods, a possible reason might be that the addition of visit-level attention weakens the contribution of time embeddings.

Risk Factors

As mentioned earlier, the factors identified by this method can be roughly divided into possible risk factors and highly associated factors, for which some pieces of evidence can be found in the literature. Besides, there were still several candidate factors proposed by our model that were seldom reported, for example, HIV infection, or we could not confirm their associations, for example, abdominal pain. One possible reason is that we only considered structured data but not textual information (ie, clinical notes); therefore, that disease or symptom may not be detailed enough to understand given only a code (ie, we know abdominal pain but do not know in which part). Furthermore, according to the AUC values of the model, the results may not be precise enough and still need to be improved. Overall, our method is completely data driven,

without any predefined candidate risk factors by experts, which is different from most studies based on regression analysis [7]. We expect that our method can provide compensational information and some new findings can be further validated by clinicians or researchers.

Error Analysis

We analyzed patient samples that are likely to be false-positives or false-negatives according to the prediction probability (as we did not require an output label but only a probability indicating the risk of each patient). One possible reason for the likely false-negatives (ground truth is case, but the predicted probability for case is quite low) is the data missing problem. For example, patient A had some respiratory symptoms such as asthma, shortness of breath, and chronic airway obstruction about 6 months before the prediction date; however, all the diagnosis codes were related to heart disease and hypertension. Therefore, it is likely that some symptoms that might serve as better indications were missing. On the other hand, one explanation for some false-positives (ground truth is control, but the predicted probability for control is quite low) is the difficulty in evaluating the severity of certain diseases or symptoms. For example, patient B had continuous respiratory symptoms such as chronic airway obstruction, but without any laboratory test values or knowledge of the drugs, it is difficult to determine whether these symptoms worsened or were already well controlled. To mitigate these factors, it is desirable to integrate more variables and background knowledge into the model in the future.

Comparison With Prior Work

One of the advantages of machine learning over statistical analysis is that it can make predictions on unseen samples [15], and it might be much easier to be deployed in real-world applications. Although many studies have focused on asthma exacerbation prediction, the majority of them belong to statistics, as they did not test their model on held-out data sets [68,69]. Among other machine learning studies, multiple conventional models have been explored, such as classification and regression tree [11,70], random forest [71], LR [72], and support vector machines [16]. However, none of these previous studies used deep learning as we know. Compared with these conventional machine learning methods, our deep learning-based method has multiple advantages. First, no feature engineering is needed, which will extremely reduce the laborious cost and expertise at the first step, for example, in comparison, Luo et al [20] included 235 features designed by multiple clinical experts as inputs that might cost a lot, but we input all the clinical codes to the model and kept their original formats without any feature selection. Second, LSTM structures can integrate any temporal patterns; thus, dependencies between variables can be easily modeled. Third, deep learning methods usually obtain better performances compared with conventional machine learning methods because of their capacities in modeling complex data structures [22,23,27], which was also proved in our experiments (compared with LR).

Furthermore, compared with previous studies, we are the first to make comprehensive visualization and personalization over the associated factors. One paper mentioned *personalize*, but it

only discussed it as a future possibility [16]. In comparison, our method showed not only cohort-level factors but also temporal-based personalized risks and factors, which would greatly facilitate precise medicine. Meanwhile, as we did not limit our input to the predefined factors, we were able to find new potential risk factors. However, one drawback of deep learning-based methods compared with the previous shallow methods is the lack of interpretability from some perspectives, for example, they can hardly report statistical evaluation measures such as *P* values and CIs, which might need further exploration.

Limitations and Future Work

Using deep learning, we offered a novel means of identifying possible risk factors and predicting the risk of asthma exacerbation. However, this study has some limitations. First, for the model interpretation part, how multiple clinical variables interact with each other needs further exploration; simply considering each variable independently but ignoring the dependency patterns between them might be insufficient for interpretation, for example, the prescription of a drug might be closely associated with a disease or symptom. Second, structured EHRs have their own drawbacks, such as data irregularity, sparsity, and noise. Thus, some potential risk factors for asthma exacerbations might not be recorded or might even be incorrectly recorded in EHRs. As a result, information integrity cannot be

guaranteed. We may need to find ways to make the data complete and more reliable, such as including information from textual reports or patient surveys. Third, it is still difficult for computer programs alone to distinguish between asthma symptoms and risk factors, and knowledge injection is needed in the future. Finally, the performance of the model still has room for improvement. It might be boosted further by designing more powerful structures or including background knowledge.

Conclusions

In this paper, we proposed an attentive deep learning-based model for asthma exacerbation prediction and employed elapsed time embeddings to model the time decays. By leveraging the weights of the model, we not only generated personalized heatmaps and specific risk scores at the individual level but also identified possible risk factors at the cohort level. Compared with previous studies, our model is effective in modeling time information and obtains better overall AUCs. As the model is completely data driven and relies little on feature engineering, it can easily be generalized to other prediction tasks. To the best of our knowledge, this is the first study to predict asthma exacerbation risks using a deep learning model that includes elapsed time embeddings. Some of the top-ranked risk factors identified have gained supporting evidence from previous medical studies, which proved that our method has good reliability and accuracy.

Acknowledgments

The authors thank Dr Irmgard Willcockson for proofreading. In addition, the authors thank Cerner for providing the valuable Health Facts electronic medical record data. The authors gratefully acknowledge the support of Nvidia Corporation with the donation of the Quadro P6000 and Titan XP GPUs used for this research.

This research was partially supported by the National Library of Medicine of the National Institutes of Health under award number R01LM011829, the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number 1R01AI130460, National Center for Advancing Translational Sciences of the National Institutes of Health under award number U01TR02062, and the Cancer Prevention Research Institute of Texas Training Grant #RP160015.

Authors' Contributions

CT conceived the research project. YX, HJ, and CT designed the pipeline and method. YX implemented the deep learning model of the study and prepared the manuscript. HJ completed the clinical part of the manuscript. WZ and HX provided valuable suggestions on cohort selection and experimental design. Y Zhou and Y Zhang extracted, cleaned the data, and performed statistics. LR helped to reorganize the data and performed normalizations for the revised version. FL, JD, SW, DZ, and CT proofread the paper and provided valuable suggestions. All the authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary materials.

[[DOCX File , 145 KB-Multimedia Appendix 1](#)]

References

1. Asthma. World Health Organization. 2017. URL: <http://www.who.int/mediacentre/factsheets/fs307/en/> [accessed 2020-06-29]
2. Most Recent Asthma Data. Center for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/asthma/most_recent_data.htm [accessed 2020-06-29]
3. Nurmagametov T, Kuwahara R, Garbe P. The economic burden of asthma in the United States, 2008-2013. *Ann Am Thorac Soc* 2018 Mar;15(3):348-356. [doi: [10.1513/AnnalsATS.201703-259OC](https://doi.org/10.1513/AnnalsATS.201703-259OC)] [Medline: [29323930](https://pubmed.ncbi.nlm.nih.gov/29323930/)]

4. Pocket Guide for Asthma Management and Prevention. Global Initiative for Asthma. 2018. URL: https://ginasthma.org/wp-content/uploads/2018/03/wms-GINA-main-pocket-guide_2018-v1.0.pdf [accessed 2020-06-29]
5. Wark PA, Gibson PG. Asthma exacerbations 3: pathogenesis. *Thorax* 2006 Oct;61(10):909-915 [FREE Full text] [doi: [10.1136/thx.2005.045187](https://doi.org/10.1136/thx.2005.045187)] [Medline: [17008482](https://pubmed.ncbi.nlm.nih.gov/17008482/)]
6. Levy ML. The national review of asthma deaths: what did we learn and what needs to change? *Breathe (Sheff)* 2015 Mar;11(1):14-24 [FREE Full text] [doi: [10.1183/20734735.008914](https://doi.org/10.1183/20734735.008914)] [Medline: [26306100](https://pubmed.ncbi.nlm.nih.gov/26306100/)]
7. Fleming L. Asthma exacerbation prediction: recent insights. *Curr Opin Allergy Clin Immunol* 2018 Apr;18(2):117-123. [doi: [10.1097/ACI.0000000000000428](https://doi.org/10.1097/ACI.0000000000000428)] [Medline: [29406359](https://pubmed.ncbi.nlm.nih.gov/29406359/)]
8. Azizpour Y, Delpisheh A, Montazeri Z, Sayehmiri K, Darabi B. Effect of childhood BMI on asthma: a systematic review and meta-analysis of case-control studies. *BMC Pediatr* 2018 Apr 26;18(1):143 [FREE Full text] [doi: [10.1186/s12887-018-1093-z](https://doi.org/10.1186/s12887-018-1093-z)] [Medline: [29699517](https://pubmed.ncbi.nlm.nih.gov/29699517/)]
9. Martin A, Bauer V, Datta A, Masi C, Mosnaim G, Solomonides A, et al. Development and validation of an asthma exacerbation prediction model using electronic health record (EHR) data. *J Asthma* 2019 Aug 8:1-8. [doi: [10.1080/02770903.2019.1648505](https://doi.org/10.1080/02770903.2019.1648505)] [Medline: [31340688](https://pubmed.ncbi.nlm.nih.gov/31340688/)]
10. Kimura H, Konno S, Makita H, Taniguchi N, Shimizu K, Suzuki M, Hi-CARAT investigators. Prospective predictors of exacerbation status in severe asthma over a 3-year follow-up. *Clin Exp Allergy* 2018 Sep;48(9):1137-1146. [doi: [10.1111/cea.13170](https://doi.org/10.1111/cea.13170)] [Medline: [29781543](https://pubmed.ncbi.nlm.nih.gov/29781543/)]
11. Lieu TA, Quesenberry CP, Sorel ME, Mendoza GR, Leong AB. Computer-based models to identify high-risk children with asthma. *Am J Respir Crit Care Med* 1998 Apr;157(4 Pt 1):1173-1180. [doi: [10.1164/ajrccm.157.4.9708124](https://doi.org/10.1164/ajrccm.157.4.9708124)] [Medline: [9563736](https://pubmed.ncbi.nlm.nih.gov/9563736/)]
12. Stanford RH, Nagar S, Lin X, O'Connor RD. Use of ICS/LABA on asthma exacerbation risk in patients within a medical group. *J Manag Care Spec Pharm* 2015 Nov;21(11):1014-1019 [FREE Full text] [doi: [10.18553/jmcp.2015.21.11.1014](https://doi.org/10.18553/jmcp.2015.21.11.1014)] [Medline: [26521113](https://pubmed.ncbi.nlm.nih.gov/26521113/)]
13. Blaiss MS, Steven GC, Bender B, Bukstein DA, Meltzer EO, Winders T. Shared decision making for the allergist. *Ann Allergy Asthma Immunol* 2019 May;122(5):463-470 [FREE Full text] [doi: [10.1016/j.anai.2018.08.019](https://doi.org/10.1016/j.anai.2018.08.019)] [Medline: [30201469](https://pubmed.ncbi.nlm.nih.gov/30201469/)]
14. Loymans RJ, Debray TP, Honkoop PJ, Termeer EH, Snoeck-Stroband JB, Schermer TR, et al. Exacerbations in adults with asthma: a systematic review and external validation of prediction models. *J Allergy Clin Immunol Pract* 2018;6(6):1942-52.e15. [doi: [10.1016/j.jaip.2018.02.004](https://doi.org/10.1016/j.jaip.2018.02.004)] [Medline: [29454163](https://pubmed.ncbi.nlm.nih.gov/29454163/)]
15. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods* 2018 Apr;15(4):233-234 [FREE Full text] [doi: [10.1038/nmeth.4642](https://doi.org/10.1038/nmeth.4642)] [Medline: [30100822](https://pubmed.ncbi.nlm.nih.gov/30100822/)]
16. Finkelstein J, Jeong IC. Machine learning approaches to personalize early prediction of asthma exacerbations. *Ann N Y Acad Sci* 2017 Jan;1387(1):153-165 [FREE Full text] [doi: [10.1111/nyas.13218](https://doi.org/10.1111/nyas.13218)] [Medline: [27627195](https://pubmed.ncbi.nlm.nih.gov/27627195/)]
17. Dexheimer J, Brown L, Leegon J, Aronsky D. Comparing decision support methodologies for identifying asthma exacerbations. *Stud Health Technol Inform* 2007;129(Pt 2):880-884. [Medline: [17911842](https://pubmed.ncbi.nlm.nih.gov/17911842/)]
18. Sanders D, Aronsky D. Detecting asthma exacerbations in a pediatric emergency department using a Bayesian network. *AMIA Annu Symp Proc* 2006:684-688 [FREE Full text] [Medline: [17238428](https://pubmed.ncbi.nlm.nih.gov/17238428/)]
19. Spyroglou II, Spöck G, Rigas AG, Paraskakis EN. Evaluation of Bayesian classifiers in asthma exacerbation prediction after medication discontinuation. *BMC Res Notes* 2018 Jul 31;11(1):522 [FREE Full text] [doi: [10.1186/s13104-018-3621-1](https://doi.org/10.1186/s13104-018-3621-1)] [Medline: [30064478](https://pubmed.ncbi.nlm.nih.gov/30064478/)]
20. Luo G, He S, Stone BL, Nkoy FL, Johnson MD. Developing a model to predict hospital encounters for asthma in asthmatic patients: secondary analysis. *JMIR Med Inform* 2020 Jan 21;8(1):e16080 [FREE Full text] [doi: [10.2196/16080](https://doi.org/10.2196/16080)] [Medline: [31961332](https://pubmed.ncbi.nlm.nih.gov/31961332/)]
21. Bae S, Choi I, Kim N. Acoustic Scene Classification Using Parallel Combination of LSTM and CNN. In: *Detection and Classification of Acoustic Scenes and Events*. 2016 Presented at: DCASE'16; September 3, 2016; Budapest, Hungary URL: http://www.cs.tut.fi/sgn/arg/dcase2016/documents/challenge_technical_reports/DCASE2016_Bae_1025.pdf
22. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015 May 28;521(7553):436-444. [doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539)] [Medline: [26017442](https://pubmed.ncbi.nlm.nih.gov/26017442/)]
23. Baytas I, Xiao C, Zhang X, Wang F, Jain A, Zhou J. Patient Subtyping via Time-Aware LSTM Networks. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017 Presented at: KDD'17; August 13-17, 2017; Halifax, NS, Canada. [doi: [10.1145/3097983.3097997](https://doi.org/10.1145/3097983.3097997)]
24. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *J Am Med Inform Assoc* 2018 Oct 1;25(10):1419-1428 [FREE Full text] [doi: [10.1093/jamia/ocy068](https://doi.org/10.1093/jamia/ocy068)] [Medline: [29893864](https://pubmed.ncbi.nlm.nih.gov/29893864/)]
25. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018;1:18 [FREE Full text] [doi: [10.1038/s41746-018-0029-1](https://doi.org/10.1038/s41746-018-0029-1)] [Medline: [31304302](https://pubmed.ncbi.nlm.nih.gov/31304302/)]
26. Ma F, Chitta R, Zhou J, You Q, Sun T, Gao J. Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017 Presented at: KDD'17; August 13-17, 2017; Halifax, NS, Canada. [doi: [10.1145/3097983.3098088](https://doi.org/10.1145/3097983.3098088)]

27. Choi E, Bahadori M, Sun J, Kulas J, Schuetz A, Stewart W. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. ArXiv 2016:- epub ahead of print(1608.05745) [[FREE Full text](#)]
28. Guo W, Ge W, Cui L, Li H, Kong L. An interpretable disease onset predictive model using crossover attention mechanism from electronic health records. IEEE Access 2019;7:134236-134244. [doi: [10.1109/access.2019.2928579](#)]
29. Xu Y, Biswal S, Deshpande S, Maher K, Sun J. RAIM: Recurrent Attentive and Intensive Model of Multimodal Patient Monitoring Data. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018 Presented at: KDD'18; August 19-23, 2018; London, UK. [doi: [10.1145/3219819.3220051](#)]
30. Wu S, Liu S, Sohn S, Moon S, Wi C, Juhn Y, et al. Modeling asynchronous event sequences with RNNs. J Biomed Inform 2018 Jul;83:167-177 [[FREE Full text](#)] [doi: [10.1016/j.jbi.2018.05.016](#)] [Medline: [29883623](#)]
31. Huang Y, Yang X, Xu C. Time-Guided High-Order Attention Model of Longitudinal Heterogeneous Healthcare Data. In: Pacific Rim International Conference on Artificial Intelligence. 2019 Presented at: PRICAI'19; August 23-30, 2019; Cuvu, Yanuka Island, Fiji. [doi: [10.1007/978-3-030-29908-8_5](#)]
32. Zhang Y, Yang X, Ivy J, Chi M. ATTAIN: Attention-based Time-Aware LSTM Networks for Disease Progression Modeling. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. 2019 Presented at: IJCAI'19; August 10-16, 2019; Vienna, Austria. [doi: [10.24963/ijcai.2019/607](#)]
33. Graves A, Mohamed A, Hinton G. Speech Recognition With Deep Recurrent Neural Networks. In: IEEE International Conference on Acoustics, Speech and Signal Processing. 2013 Presented at: ICASSP'13; May 26-31, 2013; Vancouver, BC, Canada. [doi: [10.1109/icassp.2013.6638947](#)]
34. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. Comput Biol Med 2017 Oct 1;89:248-255. [doi: [10.1016/j.combiomed.2017.08.015](#)] [Medline: [28843829](#)]
35. Do QT, Doig AK, Son TC. Deep Q-learning for predicting asthma attack with considering personalized environmental triggers' risk scores. Conf Proc IEEE Eng Med Biol Soc 2019 Jul;2019:562-565. [doi: [10.1109/EMBC.2019.8857172](#)] [Medline: [31945961](#)]
36. Vaswani A, Shazeer N, Parmar N. Attention Is All You Need. NIPS Proceedings. 2017. URL: <https://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf> [accessed 2020-07-14]
37. Zeng D, Liu K, Lai S, Zhou G, Zhao J. Relation Classification via Convolutional Deep Neural Network. In: 25th International Conference on Computational Linguistics: Technical Papers. 2014 Presented at: COLING'14; August 23-29, 2014; Dublin, Ireland.
38. Amazon Simple Storage Service Endpoints and Quotas - AWS. URL: <https://s3-us-west-2> [accessed 2020-06-29]
39. Song H, Rajan D, Thiagarajan J, Spanias A. Attend and diagnose: clinical time series analysis using attention models. ArXiv 2018:- epub ahead of print [[FREE Full text](#)]
40. Nair V, Hinton G. Rectified Linear Units Improve Restricted Boltzmann Machines. In: Proceedings of the 27th International Conference on Machine Learning. 2010 Presented at: ICML'10; June 21-24, 2010; Haifa, Israel URL: <https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf>
41. Xu B, Wang N, Chen T, Li M. Empirical evaluation of rectified activations in convolutional network. ArXiv 2015:- epub ahead of print [[FREE Full text](#)]
42. Hinton G, Srivastava N, Swersky K. Neural networks for machine learning. Coursera, video lectures 2012;264:1. [doi: [10.1002/da.20889](#)] [Medline: [22307921](#)]
43. Kingma D, Ba J. Adam: a method for stochastic optimization. ArXiv 2014:- epub ahead of print [[FREE Full text](#)]
44. Abadi M, Agarwal A, Barham P. Tensorflow: large-scale machine learning on heterogeneous distributed systems. ArXiv 2016:- epub ahead of print [[FREE Full text](#)]
45. Ioffe S, Szegedy C. Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32 nd International Conference on Machine Learning. URL: <http://proceedings.mlr.press/v37/ioffe15.pdf> [accessed 2020-07-14]
46. Ghosh N, Choudhury P, Subramani E, Saha D, Sengupta S, Joshi M, et al. Metabolomic signatures of asthma-COPD overlap (ACO) are different from asthma and COPD. Metabolomics 2019 Jun 04;15(6):87. [doi: [10.1007/s11306-019-1552-z](#)] [Medline: [31165288](#)]
47. Naik RD, Vaezi MF. Extra-esophageal gastroesophageal reflux disease and asthma: understanding this interplay. Expert Rev Gastroenterol Hepatol 2015 Jul;9(7):969-982. [doi: [10.1586/17474124.2015.1042861](#)] [Medline: [26067887](#)]
48. Solidoro P, Patrucco F, Fagoonee S, Pellicano R. Asthma and gastroesophageal reflux disease: a multidisciplinary point of view. Minerva Med 2017 Aug;108(4):350-356. [doi: [10.23736/S0026-4806.17.05181-3](#)] [Medline: [28378957](#)]
49. Peng YH, Chen KF, Liao WC, Hsia TC, Chen HJ, Yin MC, et al. Association of migraine with asthma risk: a retrospective population-based cohort study. Clin Respir J 2018 Mar;12(3):1030-1037. [doi: [10.1111/crj.12623](#)] [Medline: [28268255](#)]
50. Dirican N, Demirci S, Cakir M. The relationship between migraine headache and asthma features. Acta Neurol Belg 2017 Jun;117(2):531-536. [doi: [10.1007/s13760-017-0764-0](#)] [Medline: [28258562](#)]
51. Savelloni J, Gunter H, Lee KC, Hsu C, Yi C, Edmonds KP, et al. Risk of respiratory depression with opioids and concomitant gabapentinoids. J Pain Res 2017;10:2635-2641 [[FREE Full text](#)] [doi: [10.2147/JPR.S144963](#)] [Medline: [29180889](#)]
52. Parmar MS. Exacerbation of asthma secondary to fentanyl transdermal patch. BMJ Case Rep 2009;2009:- [[FREE Full text](#)] [doi: [10.1136/bcr.10.2008.1062](#)] [Medline: [21686475](#)]

53. Henderson AJ, Shaheen SO. Acetaminophen and asthma. *Paediatr Respir Rev* 2013 Mar;14(1):9-15; quiz 16. [doi: [10.1016/j.prrv.2012.04.004](https://doi.org/10.1016/j.prrv.2012.04.004)] [Medline: [23347656](https://pubmed.ncbi.nlm.nih.gov/23347656/)]
54. Web MD. URL: <https://www.webmd.com/asthma/news/20091105/acetaminophen-may-be-linked-to-asthma-risk#1>
55. Morales DR, Guthrie B, Lipworth BJ, Jackson C, Donnan PT, Santiago VH. NSAID-exacerbated respiratory disease: a meta-analysis evaluating prevalence, mean provocative dose of aspirin and increased asthma morbidity. *Allergy* 2015 Jul;70(7):828-835. [doi: [10.1111/all.12629](https://doi.org/10.1111/all.12629)] [Medline: [25855099](https://pubmed.ncbi.nlm.nih.gov/25855099/)]
56. Bai TR, Vonk JM, Postma DS, Boezen HM. Severe exacerbations predict excess lung function decline in asthma. *Eur Respir J* 2007 Sep;30(3):452-456 [FREE Full text] [doi: [10.1183/09031936.00165106](https://doi.org/10.1183/09031936.00165106)] [Medline: [17537763](https://pubmed.ncbi.nlm.nih.gov/17537763/)]
57. Mapping Between ICD-10 and ICD-9. Ministry of Health NZ. URL: <https://www.health.govt.nz/nz-health-statistics/data-references/mapping-tools/mapping-between-icd-10-and-icd-9> [accessed 2019-01-01]
58. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *ArXiv* 2014:- epub ahead of print [FREE Full text]
59. Xiang Y, Chen Q, Wang X, Qin Y. Answer selection in community question answering via attentive neural networks. *IEEE Signal Process Lett* 2017 Apr;24(4):505-509. [doi: [10.1109/lsp.2017.2673123](https://doi.org/10.1109/lsp.2017.2673123)]
60. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical Attention Networks for Document Classification. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies at: Proceedings of the. 2016 Presented at: ACL'16; June 12-17, 2016; San Diego, California.* [doi: [10.18653/v1/n16-1174](https://doi.org/10.18653/v1/n16-1174)]
61. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997 Nov 15;9(8):1735-1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)] [Medline: [9377276](https://pubmed.ncbi.nlm.nih.gov/9377276/)]
62. Zia T, Zahid U. Long short-term memory recurrent neural network architectures for Urdu acoustic modeling. *Int J Speech Technol* 2018 Nov 8;22(1):21-30. [doi: [10.1007/s10772-018-09573-7](https://doi.org/10.1007/s10772-018-09573-7)]
63. Mandic S, Go C, Aggarwal I, Myers J, Froelicher VF. Relationship of predictive modeling to receiver operating characteristics. *J Cardiopulm Rehabil* 2008;28(6):415-419. [doi: [10.1097/hcr.0b013e3181818c3c78](https://doi.org/10.1097/hcr.0b013e3181818c3c78)]
64. Hosmer JD, Lemeshow S, Sturdivant R. *Applied Logistic Regression*. New York, USA: John Wiley & Sons; 2013.
65. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002 Jun 1;16:321-357. [doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953)]
66. Pal S, Mitra S. Multilayer perceptron, fuzzy sets, and classification. *IEEE Trans Neural Netw* 1992;3(5):683-697. [doi: [10.1109/72.159058](https://doi.org/10.1109/72.159058)] [Medline: [18276468](https://pubmed.ncbi.nlm.nih.gov/18276468/)]
67. GitHub. URL: https://github.com/o0windseed00/asthma_prediction [accessed 2020-07-14]
68. Sato R, Tomita K, Sano H, Ichihashi H, Yamagata S, Sano A, et al. The strategy for predicting future exacerbation of asthma using a combination of the asthma control test and lung function test. *J Asthma* 2009 Sep;46(7):677-682. [doi: [10.1080/02770900902972160](https://doi.org/10.1080/02770900902972160)] [Medline: [19728204](https://pubmed.ncbi.nlm.nih.gov/19728204/)]
69. Eisner MD, Yegin A, Trzaskoma B. Severity of asthma score predicts clinical outcomes in patients with moderate to severe persistent asthma. *Chest* 2012 Jan;141(1):58-65. [doi: [10.1378/chest.11-0020](https://doi.org/10.1378/chest.11-0020)] [Medline: [21885725](https://pubmed.ncbi.nlm.nih.gov/21885725/)]
70. Lieu TA, Capra AM, Quesenberry CP, Mendoza GR, Mazar M. Computer-based models to identify high-risk adults with asthma: is the glass half empty of half full? *J Asthma* 1999 Jun;36(4):359-370. [doi: [10.3109/02770909909068229](https://doi.org/10.3109/02770909909068229)] [Medline: [10386500](https://pubmed.ncbi.nlm.nih.gov/10386500/)]
71. Xu M, Tantisira KG, Wu A, Litonjua AA, Chu J, Himes BE, et al. Genome wide association study to predict severe asthma exacerbations in children using random forests classifiers. *BMC Med Genet* 2011 Jun 30;12:90 [FREE Full text] [doi: [10.1186/1471-2350-12-90](https://doi.org/10.1186/1471-2350-12-90)] [Medline: [21718536](https://pubmed.ncbi.nlm.nih.gov/21718536/)]
72. Loymans RJ, Honkoop PJ, Termeer EH, Snoeck-Stroband JB, Assendelft WJ, Schermer TRJ, et al. Identifying patients at risk for severe exacerbations of asthma: development and external validation of a multivariable prediction model. *Thorax* 2016 Sep;71(9):838-846. [doi: [10.1136/thoraxjnl-2015-208138](https://doi.org/10.1136/thoraxjnl-2015-208138)] [Medline: [27044486](https://pubmed.ncbi.nlm.nih.gov/27044486/)]

Abbreviations

- ALSTM:** attention long short-term memory
- ATTAIN:** Attention-based Time-aware Disease Progression
- AUC:** area under the receiver operating curve
- EHR:** electronic health record
- ICD:** International Classification of Disease Code
- ICS:** inhaled corticosteroids
- LR:** logistic regression
- LSTM:** long short-term memory
- RETAIN:** Reverse Time Attention model
- RNN:** recurrent neural network
- SABA:** short-acting beta agonist
- TLSTM:** time-aware long short-term memory

TSANN: time-sensitive attentive neural network

Edited by G Eysenbach; submitted 09.11.19; peer-reviewed by Y Hafeez, S Chen, X Liu; comments to author 30.12.19; revised version received 02.03.20; accepted 13.05.20; published 31.07.20

Please cite as:

Xiang Y, Ji H, Zhou Y, Li F, Du J, Rasmy L, Wu S, Zheng WJ, Xu H, Zhi D, Zhang Y, Tao C

Asthma Exacerbation Prediction and Risk Factor Analysis Based on a Time-Sensitive, Attentive Neural Network: Retrospective Cohort Study

J Med Internet Res 2020;22(7):e16981

URL: <https://www.jmir.org/2020/7/e16981>

doi: [10.2196/16981](https://doi.org/10.2196/16981)

PMID: [32735224](https://pubmed.ncbi.nlm.nih.gov/32735224/)

©Yang Xiang, Hangyu Ji, Yujia Zhou, Fang Li, Jingcheng Du, Laila Rasmy, Stephen Wu, W Jim Zheng, Hua Xu, Degui Zhi, Yaoyun Zhang, Cui Tao. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 31.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.