

# Learning Latent Space Representation with Correlational Neural Network to Predict Patient Outcome using Electronic Health Records

**Subendhu Rongali**

University of Massachusetts, Amherst  
srongali@cs.umass.edu

**Hong Yu**

University of Massachusetts, Amherst  
hongyu@cs.umass.edu

## Abstract

Predicting clinical outcomes using longitudinal electronic health record (EHRs) data is a clinically important and computationally challenging task. In this study, we report CLOUT, an LSTM-based predictive model, which uses a correlational neural network model to identify a latent space representation of three clinical features - ICD codes, labs, and medications during a patient's encounter. Experiments on the MIMIC-III dataset show that CLOUT improves performance on predicting patient mortality, surpassing previous state-of-the-art models.

## 1 Introduction

High-precision predictive modeling of clinical outcomes (e.g., adverse events such as onset of a disease and death) is a computationally challenging but clinically important task. If physicians can be notified about potentially adverse events in advance, they may take steps to prevent them. Electronic health records (EHRs) are widely used in the US hospitals and have been actively researched for predictive modeling.

Neural networks have shown a great success for many artificial intelligence applications including computer vision and natural language processing [LeCun *et al.*, 1989]. Among different neural network models, recurrent neural networks [Mikolov *et al.*, 2010] are a sub-class of models that work well with temporally spaced data. There have been promising results showing the use of recurrent neural network in clinical applications [Lipton *et al.*, 2015]. Auto-encoders [Vincent *et al.*, 2010] are another class of neural networks that allow us to extract high-level representations of complex data such as EHRs using large unlabelled data.

In this work, we report how we use information from patients' EHRs to make predictions for adverse events like patient mortality using recurrent neural network models. Unlike previous works that built predictive models mainly using the structured ICD data alone, ignoring other fields that capture a significant amount of clinical information, we make use of all the available information. We built a predictive model called CLOUT (for Lstm OUTcome prediction using Comprehensive features), that integrates clinically significant information available in EHRs into an LSTM model to predict patients' health outcomes.

## 2 Related work

The MIMIC-III dataset, a publicly available EHR dataset made available by the MIT Lab for Computational Physiology, has been widely used for predictive models [Johnson *et al.*, 2016]. RETAIN [Choi *et al.*, 2016] only made use of the structured ICD-codes for predictions, ignoring other data such as medications and lab tests. Deep Patient [Miotto *et al.*, 2016] uses multiple layers of denoising autoencoders to build a high level representation of patient data. They use comprehensive patient data but ignore its temporal aspect by combining all the information across time into a single vector. Data-driven phenotyping was explored in some recent works [Zhou *et al.*, 2014] but that was again mostly used with structured diagnosis codes - like the DxGroups or HCC codes.

## 3 The MIMIC-III Dataset, Pre-Processing, and Outcome Label

We used MIMIC-III [Johnson *et al.*, 2016] for all our experiments. The data-set contains 7,537 patients with 2 or more encounters, which we used to build our LSTM models. Note that an encounter represents a single visit or admission of the patient to the hospital.

Each encounter contains the diagnosis codes documented in that encounter, the prescribed medications, the lab tests and their results (e.g., normal and abnormal), and the notes written by the physicians.

We converted the structured data into a one-hot vector of the size of the vocabulary, which is the total numbers of distinct ICD codes, labs, and medications recorded. After processing the MIMIC-III dataset, the data we used in this study includes 7,537 patients, with a total of 942 different ICD codes, 3,202 different medications, 681 different labs tests, and 284 different abnormal lab tests. We call this reduced dataset, r-MIMIC. The outcome label we used in this study is the patient mortality, obtained from the hospital records and the social security death records.

## 4 Methodology

CLOUT is the overall model indicating the use of comprehensive data in an LSTM. We implemented multiple CLOUT models for prediction, that use the data in different ways.

## 4.1 A Simple LSTM model

This CLOUT implementation takes a sequence of encounters for each patient as input. Each encounter  $i$ th is represented by three one-hot vectors for the ICD codes ( $c_i$ ), labs ( $l_i$ ) and medications ( $m_i$ ).  $m_i$  includes medication names, but ignores other related information such as dosage and route.  $l_i$  represents whether the lab test is normal or abnormal. We then pass the representation of each encounter through a linear embedding layer to get a dense vector representations. Let the weights of the corresponding linear layers be  $W_c$ ,  $W_l$ , and  $W_m$ . We now concatenate these dense vectors to obtain the final encounter vector  $e_i$ .  $e_i$  is therefore given by  $e_i = [W_c.c_i, W_l.l_i, W_m.m_i]$ .

We then pass the encounter vectors  $[e_1, e_2, \dots, e_n]$  for each patient through an LSTM. We use a linear layer and squash the output using a sigmoid function to obtain the probabilities for the positive output label for the given patient. During training, we use the binary cross entropy loss function.

## 4.2 LSTM with Latent-Space Encounter Vectors

ICD, lab and medications are not independent events or views. They are clinically intertwined or correlated. For example, the lab tests and medications depend upon the diagnoses of the patient in that particular encounter. Therefore, we built a multi-view learning CLOUT model to capture such correlations using a latent space representation vector.

Our latent space representation is independent of individual views. In this representation, projecting the ICD codes vector and lab tests vector, for example, into this latent space should give us vectors that are very close to each other. Therefore, our latent space representation captures the patterns that exist in the views. We can then use the latent space representation of each encounter in our predictions.

We use a correlational neural network described in [Chandar *et al.*, 2016] to get the latent space representations for all three views - ICD codes, medications and labs at once. We then concatenate this latent space vector to our existing information to create the final input vector.

## 5 Results and Conclusion

The AUC-ROC results for CLOUT models are significantly better compared to both RETAIN, and the LSTM with just ICD codes. We also present the classification reports.

Table 1: AUC-ROC scores

Method	AUC-ROC	Epoch
RETAIN (only ICD)	0.8225	16
LSTM (only ICD)	0.8239	1
CLOUT - Simple	0.8712	1
<b>CLOUT - Latent vectors</b>	<b>0.8854</b>	<b>1</b>

## References

[Chandar *et al.*, 2016] Sarath Chandar, Mitesh M Khapra, Hugo Larochelle, and Balaraman Ravindran. Correlational neural networks. *Neural computation*, 28(2):257–285, 2016.

Table 2: Classification Reports

Method	Class	Precision	Recall	Fscore
LSTM	0	0.79	0.81	0.80
(ICDs only)	1	0.67	0.64	0.66
	avg	0.75	0.75	0.75
CLOUT	0	0.85	0.82	0.83
(Simple)	1	0.71	0.76	0.73
	avg	0.80	0.79	0.80
CLOUT	0	0.84	0.88	0.86
(Latent vectors)	1	0.78	0.72	0.72
	avg	<b>0.82</b>	<b>0.82</b>	<b>0.82</b>

[Choi *et al.*, 2016] Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In *Advances in Neural Information Processing Systems*, pages 3504–3512, 2016.

[Johnson *et al.*, 2016] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghaseemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3:160035, 2016.

[LeCun *et al.*, 1989] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[Lipton *et al.*, 2015] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzel. Learning to diagnose with lstm recurrent neural networks. *arXiv preprint arXiv:1511.03677*, 2015.

[Mikolov *et al.*, 2010] Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černocký, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[Miotto *et al.*, 2016] Riccardo Miotto, Li Li, Brian A Kidd, and Joel T Dudley. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports*, 6:26094, 2016.

[Vincent *et al.*, 2010] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408, 2010.

[Zhou *et al.*, 2014] Jiayu Zhou, Fei Wang, Jianying Hu, and Jieping Ye. From micro to macro: data driven phenotyping by densification of longitudinal electronic medical records. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 135–144. ACM, 2014.