Mortality Prediction of ICU Diabetic Patients, Based on Clinical History and Clinical Notes - A Multi-Network Deep Learning Model

Olabisi Balogun(olb2@illinois.edu), Jorge Flores ([jorgedf2@illinois.edu](mailto:jorgedf2@illinois.edu)), Francisco Noya ([fnoya2@illinois.edu](mailto:fnoya2@illinois.edu))

**Abstract:** Diabetes is a very serious condition that can lead to serious consequences and increase the risk of death when a patient is admitted in the ICU. In this paper, we explore the possibility of using clinical notes as well as lab and chart events in a multi-network, model where we process the clinical notes in a time series RNN with dual-attention network, and the lab and chart events in another RNN, also with multi-attention, then merge them into a single multi-network model to try to predict mortality of diabetic patients. Our model achieved an AUC ROC score of 0.977.

**Abbreviations**: CNN – Convolutional Neural Network, UMLS – Unified Medical Language System, ICU – Intensive Care Unit, RNN - Recurrent Neural Networks, ICD, GRU – Gated Recurrent Unit.

# Introduction

Diabetes mellitus comprises various medical conditions that are associated with hyperglycemia and it can be caused by an abnormal secretion of insulin in the body system (Egan & Dinneen, 2019). The number of diabetic patients worldwide has soared in the last 20 years (Zimmet, Magliano, Herman, & Shaw, 2014). This increment has led to different research on the efficient management of this condition. Diabetes has contributed to a 75% increase in mortality rate, a diabetic adult is at higher risk of death from other diseases compared to a non-diabetic adult in the US (Gregg et al., 2018), therefore, an efficient and timely interventional clinical decision support system that would aid in preventing death or predicting mortality would lead to a major milestone in diabetes management research. Anand et al., (2018) stated that having diabetes could affect the treatment of ICU patients as unique factors have to be taken into consideration; for example, glucose levels. Diabetic patients account for more than 45% of ICU stays and consume more resources compared to patients suffering from other chronic diseases (Anand et al., 2018). Our research study aims to build a multi-network deep learning model to predict the mortality of ICU diabetic patients using the clinical notes, and features from the Apache II scoring system plus features such as glycosylated hemoglobin (HbA1c), serum creatinine, and glucose levels, that can improve the predictability for diabetic patients (Anand et al., 2018).

# Previous Work

Research in diabetes is broad and different researches have engaged machine learning algorithms to improve the treatment and management of this global metabolic disorder. We streamlined past works to those that interest us most that were done in the past five years and targeted a prediction task.

Anand et al., (2018) employed binomial logistic regression to predict the risk mortality of ICU diabetic patients. Their research was based on the certainty that these variables (HbA1c, mean glucose during stay, serum creatine, diagnoses upon admission, and type of admission) were the major predictors needed for the mortality prediction. Their research used the data from the MIMIC-III database and their model achieved AUC values of 0.787.

Ye, Yao, Shen, Janarthanam, & Luo (2020) employed different machine learning algorithms and knowledge-guided feature extraction to predict mortality in critically ill diabetic patients. Knowledge-guided CNN using CUI (UMLS concept unique identifiers) plus word embedding and CNN using word embedding were applied to clinical notes to predict mortality in diabetic ICU patients. They also ran different machine learning models such as Logistic Regression, Random Forest, XGBoost, Gradient Boosting, Deep Learning ANN, and Majority Voting with Majority Voting model taking the lead with an AUC of 0.87. These machine learning algorithms were used with structured EHR data to predict mortality risk in ICU diabetic patients. CNN with word embedding performed best overall with AUC of 0.97.

Yang, Kuang, & Xia (2021) proposed a multimodal deep learning neural network, which uses time series data and clinical notes to predict mortality of ICU patients, who obtained an AUC ROC score of 0.861.

Che, Purushotham, Cho, Sontag, & Liu (2018) built a model – GRU\_D based on Gated Recurrent Unit (GRU), which could handle missing data, for mortality prediction using the time series features including input events, output events, lab events, and prescription events from the MIMICIII dataset.

Choi, Bahadori, Kulas, Schuetz, Stewart and Sun explore the use of multi RNNs with multi attention at different level of granularity on “RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism”, we are going to apply a similar mechanism on each of the different networks, notes and events.

# Approach and implementation

## Problem formulation:

The EHR information of a patient that is admitted into ICU can be represented as a timeseries of multiclass, multivariate observations, in our case, we will pay attention to two classes, chart/lab events and clinical notes. Considering *v* different variables for events, and an aggregate note embedding *n* for each time *t*i, of a total of *T* different distinct dates, we can represent a patient *p* (of a total of *P* patients) of a as a sequence of touples (*t*i(p), *x*i(p), *n*i (p)) where *i* = 1, …, T(p). The timestamps *t*i(p), denote the *i-*th distinct date of the *p*-th patient, and T(p), the number of distinct dates for the *p*-th patient. The goal of our network is to predict the labels at the end of the sequence y ϵ {0,1}, where 0 means the patient is predicted to be alive 48 hrs after the observation window, and 1 is predicted to be deceased.

## Data Source

For our project, we decided to use the data from MIMIC-III database, because this is a very complete source with a lot of good real-life records of patients who have been admitted to ICU.

## Cohort Selection

First, we defined our cohort as patients who have a diabetic diagnosis, that is, ICD-9 codes containing the word ‘diabetes’ exempting codes 3572, V771, V180, V1221. We considered all the data of each patient from the moment they were first admitted, all the way to 48 hrs before last discharge. Some patients had multiple admissions, so we chose the observation window to be the earliest admission time to 48 hrs before the last discharge time. Any patient who did not have any data 48 hours before final discharge was removed from the analysis. We used the death time presence to create a mortality flag, zero for alive and one for dead.

## Data Processing

The MIMIC-III database is available for access via AWS without need to download. We created a cohort master table (diabetes\_patient\_cohort), which we later used to create the source of information for each of our networks, notes and events. Using SQL Queries, we created tables such as

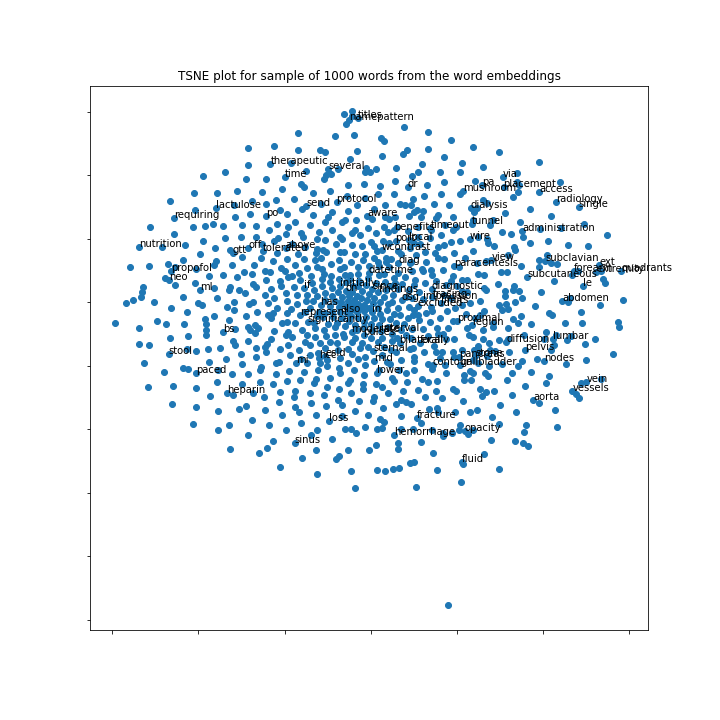
- *Diabetic\_patients\_notes\_agg* contains the notes of a patient aggregated by distinct date by concatenating the notes for that date if there are more than one.

*- Diabetic\_patients\_events* contains the actual values for the different laboratory tests, vital signs and other values from ICU charts that we selected as features. The data is aggregated by patient and date. If multiple readings were obtained for the same feature of a patient on the same date, the daily average of it was used.

We split the *diabetic\_patient\_cohort* in train and evaluation sets using a split ratio of 80% to 20%, we call this our *“unbalanced”* cohort. We observed that the data has an inherent class imbalance. To resolve this issue, we oversampled the minority class in the training set to create a *“balanced”* cohort. All these actions were done using the sklearn library, class *resample* and stored on S3 bucket to be built in the AWS Athena as tables (*train\_cohort* and *test\_cohort*).

## Clinical Notes Embedding

We formed a notes table for the diabetic patients in the ICU by aggregating the notes for a given date. Then, we preprocessed the notes and trained our own word embedding model using the Word2Vec class from the gensim.models package. Once we had our embedding model, we processed each *i*-th distinct date aggregated notes for every patient *p |* 1*, …,* P, *n*i (p) by calculating an embedding vector for each day notes, we achieved this by obtaining the embedding vector for each word and then add the embedding for all the words in that date’s notes. The following is a representation of the notes embedding using tSNE:



## Neural Network Model

The following diagram shows the model architecture:



## Notes Network

The Notes Network architecture consists of taking the clinical notes embedding described on the previous section and then feed it into two parallel RNN (GRUs) layers, to calculate on each its attention, on one of them we calculate the attention for each distinct date (α), and on the second one, we calculate the attention of each embedding component (β), followed by a FCL layer, and a Sigmoid Activation function, to output the embedding vectors for this network

## Events Network

A multi-hot vector using the codes of the features and their corresponding values was constructed for each patient and for each date. These multi-hot vectors were used as inputs for the Events Network. The architecture of this network comprises another pair of parallel RNNs (GRUs) with alpha (α) and (β) attention, where alpha means the attention for each distinct event date, and beta means the attention for the different features. The network finally has a FCL and a Sigmoid Activation function as the embedding vector of the events of each patient.

### Notes-Events Network

The final Model takes the output of the two previous Networks (embedding vectors output of Notes and Events network is of size 128), then concatenates them and applies two additional FCL layers with a dropout in the middle and a final Sigmoid activation function to generate the final prediction.

# Experimental Results

The following table shows the results of our model as we progressed in the construction and experimentation (always 10 epochs):

## 



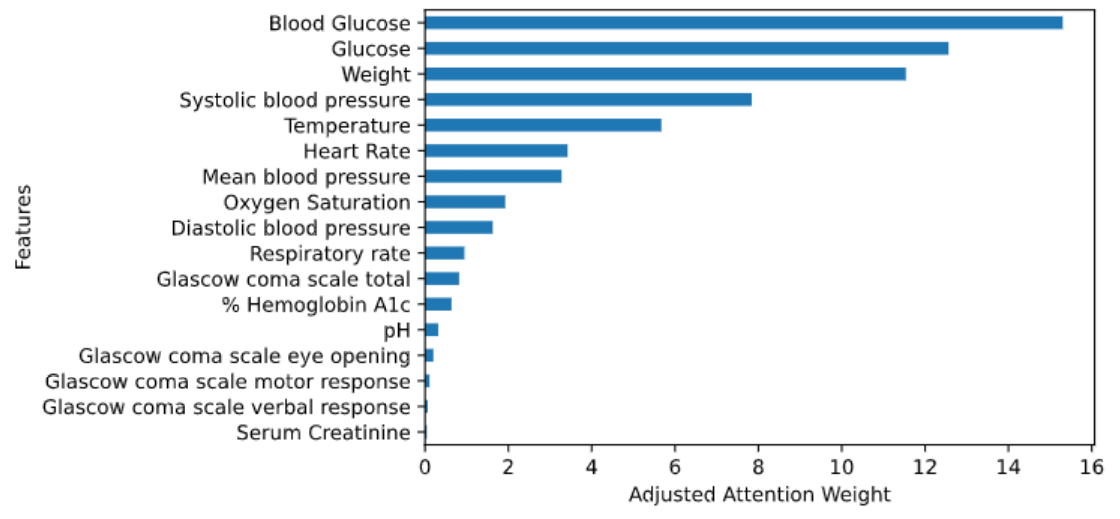
## Experimental Setup

The model was implemented in Pytorch 1.7.1 and trained initially in AWS Sagemaker environment with a memory of 8GBs. Eventually we moved the training to a more powerful machine with the same Pytorch version, but with 32GBs of RAM and an NVIDIA GeForce GTX 1060 with 6GB GPU. This enabled us to train the model and do fine tuning with much faster training times compared to only using the CPU environment.

# Discussion

We analyzed the clinical notes of diabetic patients including the lab and other clinical features collected from the MIMICIII database and fed the data into parallel RNN to predict mortality. Our full network is yielding really good results compared to other previous works, achieving a maximum AUC ROC of 0.977. We originally had observed a trade-off between the reported recall and precision when using the original training set, versus when using a balanced training set, but once we updated our models to include α and β attention on both the Events and Notes network, the balanced cohort clearly yields much better results, validating the idea of using a re-sampling algorithm to avoid our model from giving more weight to the majority class.

Upon analyzing the β attention of the events network (the weights on the different features that were considered for the Events Network, we found that the following are the features of more weight in the attention mechanism:



This seems to make absolute sense when we consider that our patients are on high risk of dead due to diabetic diagnostics.

# Conclusion/Optimization

We developed a multi-network multi-attention deep learning model to predict the mortality of diabetic ICU patients 48 hours before discharge. The evaluation of the model reported an AUC of 0.977. The model is a step further of previous works by engaging multiple networks and using both clinical notes and clinical features as source to parallel RNNs as well as implementing multi-attention on each of the parallel networks in the prediction task. This model is a step in the right direction, which could be integrated in the clinical setting to manage the extensive resources that are being expended on diabetic patients and increase the efficiency of clinicians in the ICU.

Further optimization can be explored in the future to improve the scores and validate the results found in other datasets of similar composition. Additionally, further work is required to ensure the resiliency of the model by implementing k-fold cross validation with resampling.

# References:

* Anand, R. S., Stey, P., Jain, S., Biron, D. R., Bhatt, H., Monteiro, K., ... & Chen, E. S. (2018). “Predicting mortality in diabetic ICU patients using machine learning and severity indices”. *AMIA Summits on Translational Science Proceedings*, 2018, 310.
* Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). “Recurrent neural networks for multivariate time series with missing values”. *Scientific reports*, *8*(1), 1-12.
* Egan, A. M., & Dinneen, S. F. (2019). “What is diabetes?”. *Medicine*, *47*(1), 1-4.
* Gregg, E. W., Cheng, Y. J., Srinivasan, M., Lin, J., Geiss, L. S., Albright, A. L., & Imperatore, G. (2018). “Trends in cause-specific mortality among adults with and without diagnosed diabetes in the USA: an epidemiological analysis of linked national survey and vital statistics data”. *The Lancet*, *391*(10138), 2430-2440.
* Yang, H., Kuang, L., & Xia, F. (2021). “Multimodal temporal-clinical note network for mortality prediction”. *Journal of Biomedical Semantics*, *12*(1), 1-14.
* Ye, J., Yao, L., Shen, J., Janarthanam, R., & Luo, Y. (2020). “Predicting mortality in critically ill patients with diabetes using machine learning and clinical notes”. *BMC Medical Informatics and Decision Making*, *20*(11), 1-7.
* Zimmet, P. Z., Magliano, D. J., Herman, W. H., & Shaw, J. E. (2014). “Diabetes: a 21st century challenge”. *The lancet Diabetes & endocrinology*, *2*(1), 56-64.
* Kumar, Brenai (2021). “10 Techniques to deal with Imbalanced Classes in Machine Learning”. *Analytics Vidhya*.
* Knaus, W. A., Draper, E. A., Wagner, D. P., & Zimmerman, J. E. (1985). “APACHE II: a severity of disease classification system”. *Critical care medicine*, 13(10), 818-829.
* Johnson, Alistair, et al. “MIMIC-III Clinical Database (version 1.4).”, Physionet, September 2016, <https://doi.org/10.13026/C2XW26>
* Choi, Bahadori, Kulas, Schuetz, Stewart and Sun. “RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism”. 2016