CS230

# Deep Learning for Prediction of Future Lab Testing Results through Longitudinal Lab Testing Data

**Shobha Dasari**
Department of Computer Science
Stanford University
sdasari1@@stanford.edu

## 1 Introduction

Overuse of diagnostic testing significantly contributes to high healthcare costs and may cause patients unnecessary harm, with studies shown that many diagnostic tests provide low-value results[1]. For patients who are elderly, laboratory testing is especially common. There is a strong need to minimize unnecessary, low-value lab testing in patients.

Additionally, little research has been conducted into the predictive power of using laboratory information systems data alone for the diagnosis of specific diseases. Additionally, achieving accurate predictions with laboratory data from electronic health records data would drastically reduce the computational power required to analyze all components of electronic health record data, such as clinical notes.

Most models that currently utilize patient laboratory data use machine learning [2, 3]. However, while machine learning approaches require much less training data, they also require hand-crafted features which can make them very sensitive to variation in data. Deep learning might serve as an alternative to these downsides of machine learning.

One study used a 2-layer deep learning model in conjunction with machine learning models in order to make predictions about disease diagnosis using only laboratory data, which was shown to be 92% accurate [4]. Another study attempted to predict the results of the next lab screening and evaluated different deep learning models (LSTM, CNN, M-CNN, Transformer, etc.) on longitudinal labs data for patients in the ICU in order to find values that are predicted to become abnormal in the next lab screening [5].

The goal of this algorithm is to predict future lab test results for patients, allowing for fewer necessary lab tests. The input to this algorithm is a patient's longitudinal lab testing history, with the HbA1c lab test (an indicator of diabetes) masked. We then use an LSTM model to output a predicted value for the patient's lab result at the last time step.

This study is part of a larger proof-of-concept study that seeks to use longitudinal laboratory information systems data (a currently under-analyzed source of information) to obtain valuable insights from routine laboratory testing data.

## 2 Dataset

This proprietary dataset comes from the private company Olea Health [6], which encompasses all the lab results from approximately 30% of nursing home residents in the United States, with 15-20 years of historical data. This dataset has longitudinal lab testing data for 404,569 patients. Time-series data is discretized per lab testing encounter (no standardized interval). Examples of the data are provided:

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

| itemid | hadm_id | charttime | 50801 | 50802 | 50803 | 50804 | 50805 | 50806 | 50808 | 50809 | ... | 51491 | 51492 | 51493 | 51497 | 51498 | 51501 | 51507 | 51514 | 51516 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 100009.0 | 2162-05-16T16:00:00Z | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 1 | 100009.0 | 2162-05-16T20:10:00Z | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | 5.0 | NaN | 1.0 | NaN | 1.014 | NaN | NaN | NaN | 1.0 |
| 2 | 100009.0 | 2162-05-17T01:40:00Z | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 3 | 100009.0 | 2162-05-17T12:19:00Z | NaN | 2.0 | NaN | 28.0 | NaN | 103.0 | 1.26 | 110.0 | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 4 | 100009.0 | 2162-05-17T14:43:00Z | NaN | 1.0 | NaN | 27.0 | NaN | NaN | NaN | 125.0 | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

Laboratory data commonly have many missing values and class imbalances due to the nature of collection (i.e., only select labs are collected from patients at all encounters, and most people in the population will fall under the standard result class). To account for this issue, upsampling was conducted to reflect the pre-diabetic and diabetic data at 30 times the rate than which they were present in the data.

The main challenge with this dataset is determining how to deal with missing values. According to laboratory testing domain experts, lab values are not missing at random; there may be inherent meaning when a lab value is not recorded in the data (i.e., a patient's provider might assume that a particular lab test result would be healthy and the test was not needed).

The HbA1c test is considered the gold standard laboratory test to diagnose diabetes, and the individual is considered healthy if the value is $\leq 5.6\%$, pre-diabetic if the value is between $5.7\%$ and $6.4\%$ inclusive, and diabetic if the value is $\geq 6.5\%$ [7].

## 3 Methods: Baseline

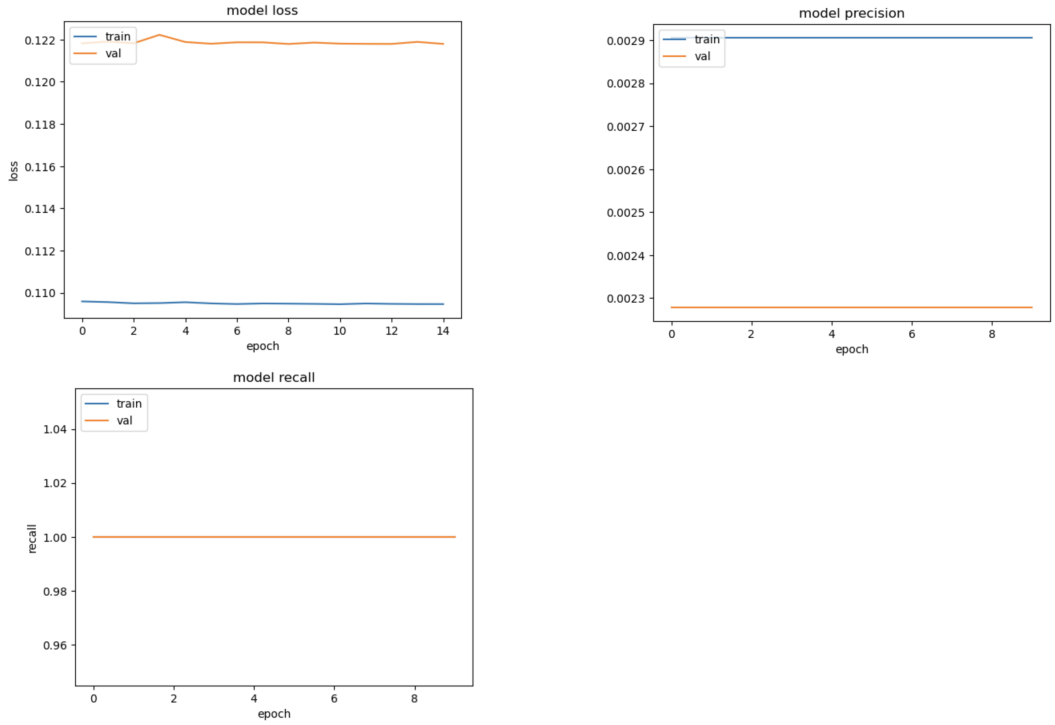
The input to this algorithm is a patient's longitudinal lab testing history. First, pre-training tasks were conducted to mask HbA1c (a predictor for diabetes) lab testing results in the data and to impute missing values with zeros.

The predictors for this model included 337 different lab testing values (though not every patient had recorded values for all lab tests). The baseline model for this was a straightforward two-layer LSTM [11] with a dense, fully-connected layer as the output layer. The model utilizes an Adam optimizer, and a categorical cross-entropy loss function. The goal of this baseline algorithm is to predict HbA1c lab results for patients based on their other lab results. We use the LSTM model to output the predicted values for the patient's HbA1C lab results at the different time steps in the data.

For this model, we evaluated loss, precision, and recall for the validation and testing sets. We did not evaluate accuracy as a metric, as it is a poor indicator of performance in a dataset with as much class imbalance as this one.

## 4 Results and Analysis: Baseline

The following graphs show the model's performance in terms of loss, precision, and recall during training.




This model's high accuracy/recall and low precision could be due to a few reasons: 1) the model discovered a lab value code that is directly correlated to HbA1C in the training data, 2) the model has severely overfit, or 3) most likely, our model is assigning the negative label of 0 (majority class) to many data examples.

The distribution of HbA1c values in the data is incredibly skewed towards 0 (representing healthy patients). 0 was also chosen as the imputed value for patients with missing values, as domain experts have advised that for patients who have missing values for HbA1c, this is likely indicative of no need to test (i.e., a healthy value). The below image shows the class distribution after upsampling the prediabetic and diabetic patient data.

```
In [8]: #get new HbA1c class counts
        labs_df[50852].value_counts()

Out[8]: 0.0    50375
        1.0     1500
        2.0     1290
        Name: 50852, dtype: int64
```

Due to this distribution, and the validation accuracy being 0.9940, I believe that the model's current performance is not completely due to overfitting. I believe that the model is predicting an HbA1c outcome of 0 for much, if not all, of the results. Possible solutions to this issue include oversampling from the data with existing HbA1c values in the training set by repeating examples in the dataset or undersampling from data without HbA1c values in the training set by eliminating examples in the dataset, until the class imbalance is reduced.

## 5 Methods: Novel Approach

The goal of this novel approach was to modify the loss function to account for class imbalance in the data, as well as account for clinical actionability of this algorithm. As most of the data falls under a healthy HbA1c value, simple categorical cross-entropy might be too punishing. In addition, misclassifying a diabetic or pre-diabetic patient is much more severe than classifying a pre-diabetic patient as diabetic. The loss function will account for this difference as well, but including a penalty matrix that assigns higher loss weights to predictions that lead to incorrect clinical actions.

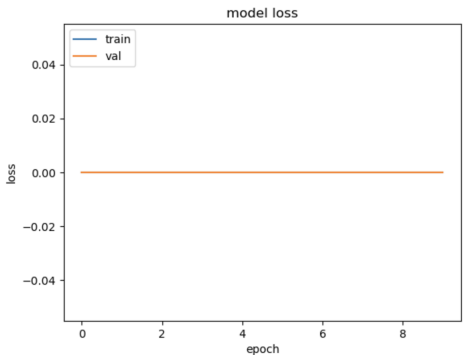
There is currently no study in published literature that utilize different custom loss functions on laboratory data, making this a unique approach to working with this data. The penalty matrix for this modified loss function is below:

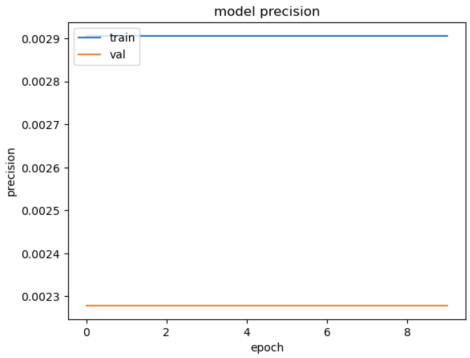
| Penalty Matrix: | y_true | | |
|---|---|---|---|
| y_pred | 0 (healthy) | 1 (pre-diabetic) | 2 (diabetic) |
| 0 (healthy) | 0.2 | 2.0 | 3.0 |
| 1 (pre-diabetic) | 2.0 | 0.5 | 1.0 |
| 2 (diabetic) | 3.0 | 1.0 | 0.5 |

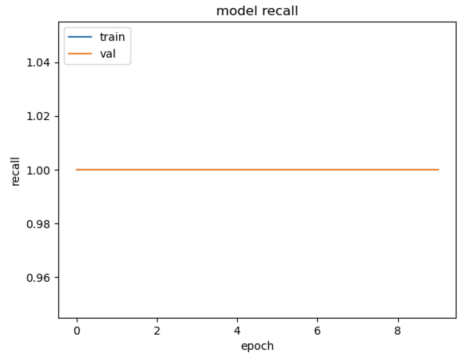
Similarly to the baseline, we evaluated loss, precision, and recall for the validation and testing sets for this model.

## 6 Results and Analysis: Novel Approach

The following graphs show the model's performance in terms of loss, precision, and recall during training.




Overall, this model was able to identify healthy patients with high recall (1.00). However, this model suffered in terms of precision (0.0023 in the validation set). This model's performance could be due to the imbalance in the dataset; it is likely that the model is assigning a label of 0 (the majority class and negative result) to many data examples.

## 7 Contributions

The analysis for this project was done by Shobha Dasari. The dataset is contributed by Olea Health.

## References


[1] Muskens JLJM, Kool RB, van Dulmen SA, et al. Overuse of diagnostic testing in healthcare: a systematic review. BMJ Quality and Safety 2022;31:54-63.

[2] Lemke, K. W., Gudzune, A. K. Kharrazi, H., Weiner, J. P. Assessing Markers From Ambulatory Laboratory Tests for Predicting High-Risk Patients. American Journal of Managed Care 2018;24(6).

[3] Tang, A., Oskotsky, T., Sirota, M. Personalizing routine lab tests with machine learning. Nat Med 27, 1514–1515 (2021). https://doi.org/10.1038/s41591-021-01486-4

[4] Park, D.J., Park, M.W., Lee, H. et al. Development of machine learning model for diagnostic disease prediction based on laboratory tests. Sci Rep 11, 2021.

[5] Ayad A, Hallawa A, Peine A, et al. Predicting Abnormalities in Laboratory Values of Patients in the Intensive Care Unit Using Different Deep Learning Models: Comparative Study. JMIR Med Inform 2022;10(8):e37658.

[6] Olea Health. "Olea Health." https://www.olea.health/. Accessed October 13, 2022.

[7] Centers for Disease Control and Prevention. "All About Your A1C." https://www.cdc.gov/diabetes/managing/managing-blood-sugar/a1c.html. Accessed December 9, 2022.