

# A Matrix Visualization system to Identify Factors Affecting Mortality Rate

Ayushi Gupta<sup>1</sup>, Veera Raghavendra Chikka<sup>1</sup> and Kamalakar Karlapalem<sup>1</sup>  
<sup>1</sup>International Institute of Information Technology, Hyderabad, Telangana, India

**Abstract** *In this paper, we present a system that visualizes the effect of parameters (such as demographic features, physical features, and medications) on mortality rate using two-dimensional matrix visualization. We take a list of conditional statements as input. These conditional statements divide the patients cohort into different sets. Size of a set is proportional to the number of patients. Comparing the size of two sets under the same constraints (conditional statements) but different expiry flag helps the user in understanding the effect of parameters on the mortality rate. We demonstrate the usefulness of our approach on a publicly available database - MIMIC-III. Our results show that the developed system can be used to validate and visualize the parameters affecting the mortality rate.*

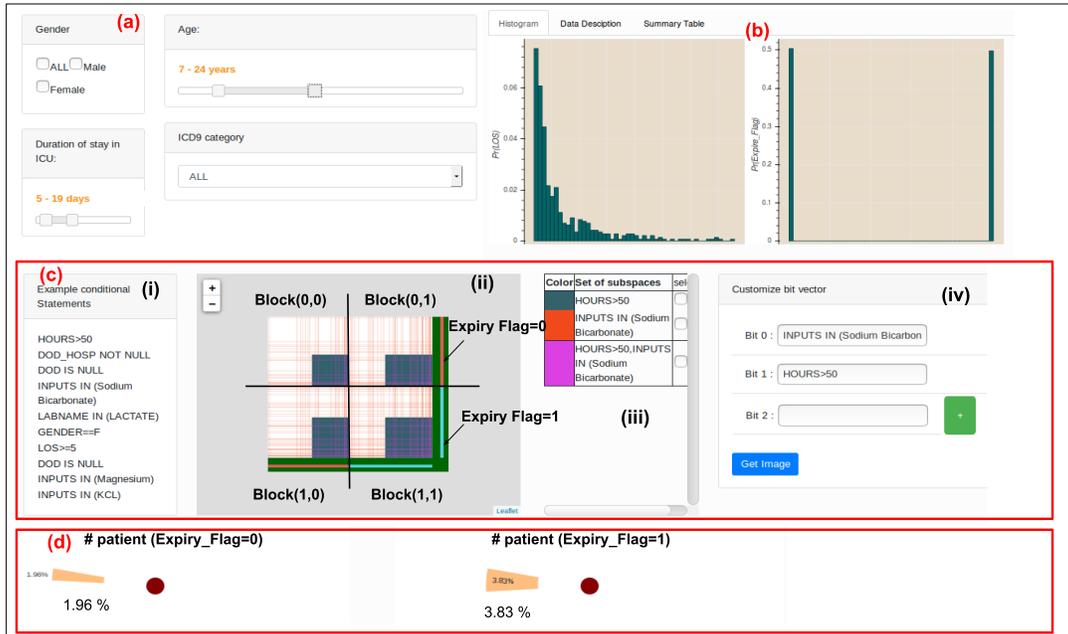
## Introduction

With the digitization of healthcare data, electronic healthcare records are growing rapidly. The healthcare data available from these systems are large both in terms of the number of records and the number of parameters for each record. The rapid availability of digitized healthcare data has increased the interest of domain experts and data analysts in understanding the information and knowledge hidden in these large datasets. The insights hidden in these large datasets are important for clinical decision support. The collective visualization of the health-related data from varying sources has the potential to reveal valuable information<sup>1</sup>. We can group patients based on various parameters such as demographic features (e.g., gender or ethnicity), physical features (e.g., weight, height or age), disease type, duration of stay in ICU or a combination of one or more parameters. Visualizing and understanding the parameters which are more susceptible to the mortality rate is the key idea of our work.

## Our methodology

In this paper, we use a publicly available dataset - MIMIC-III<sup>3</sup>. This dataset contains detailed multi-variate healthcare data of over 60,000 patients who stayed in the intensive care unit. The steps involved in our method are:

- 1. Initial Cohort Selection :** User selects the initial cohort based on International Classification of Diseases 9 (ICD9) code, Age, Length of ICU stay and Gender (Figure 1(a)).
- 2. Data Exploration :** In data exploration area, we provide a dashboard (Figure 1(b)) which visualizes the histogram plot of numeric attributes, and gives a brief statistical summary of the selected cohort (including the mean, median, minimum, maximum and standard deviation) of numeric attributes. For categorical attributes like input events, lab events and output events, we provide a top- $m$  frequent events and least- $m$  frequent events in the data description tab.
- 3. Matrix visualization :** We create a 2-D matrix of size  $N \times N$ , where  $N$  is the number of patients selected in step-1. All the patients are mapped across the row and column of this matrix. We borrow this concept from Heidi visualization<sup>2</sup>. In the matrix visualization, the order in which patients are placed across the matrix plays a significant role. We first place the patients with  $Expiry\_Flag = 0$  (alive patients) and then the patients with  $Expiry\_Flag = 1$  (dead patients) (Figure 1(c)(ii)). A user selects a set of conditional statements (Figure 1(c)(i) shows some of the illustrative conditional statements). A conditional statement gives a binary output (True or False). A  $Cell(i, j)$  in the matrix stores a bit-vector. A bit  $b_k$  in the bit vector is 1 if the result of the conditional statement  $Cond_k$  on Patients  $P_i$  and  $P_j$  is 1. We visualize this matrix as a multi-colored image. We assign a unique color to a unique bit vector. The legend (Figure 1(c)(iii)) maps the RGB color to a subset of conditional statements.
- 4. Pattern Analysis :** The matrix obtained from Step-3 is block-segmented. We number the blocks from 0 to 1 across row and column starting from the top left block. In Figure 1(c) Block  $B_{0,0}$  refers to the patients with  $Expiry\_Flag = 0$  (alive patients) and Block  $B_{1,1}$  refers to the patients with  $Expiry\_Flag = 1$  (dead patients). We consider the diagonal blocks ( $B_{0,0}$  and  $B_{1,1}$ ) to understand the parameters which are more susceptible to increased mortality rate. Selecting a color in legend triggers the gauge visualization in pattern analysis (Figure 1(d)). (Source code and other additional resources are shared at path : [https://github.com/ayushi04/visual\\_health\\_challenge](https://github.com/ayushi04/visual_health_challenge))



**Figure 1:** Matrix visualization system (Figure 1(a): Initial Cohort Selection, Figure 1(b): Data Exploration Area Figure, 1(c): Matrix visualization, Figure 1(d): Pattern Analysis Area)

### Use-case example

Consider an example scenario where the user creates an initial patient cohort by selecting all the ICU patients belonging to ICD9-category: diseases of the blood and blood-forming organs. From the data exploration area, the user finds that there are 1175 ICU patients and 576 patients are ICU survivors. In Figure 1(b) the user sees that the distribution of duration of stay in ICU is skewed and count of ICU survivors and expired patients is almost equal. The description tab gives the other additional information like top-20 frequent input events. From this list, the user observes that few input events (such as D5W, 9% Normal Saline) are given to almost half of the ICU patients. Suppose the user wants to see a correlation between the number of ICU survivors and intake of sodium bicarbonate and the duration of stay in ICU. Figure 1(c)(iv) shows the conditional statements selected for this purpose. Figure 1(c)(ii) displays the generated image and the Figure 1(c)(iii) maps the color value in image with the conditional statements. Legends at the bottom right corner of the image show that the patients with  $Expiry\_Flag = 0$  are followed by the patients with  $Expiry\_Flag = 1$ . The input image is block-segmented and the user considers the diagonal blocks ( $B_{0,0}$  and  $B_{1,1}$ ) to visualize and identify the effect of sodium bicarbonate on the mortality rate. On click on the red color in the color legend (Figure 1(c)(iii)), the user gets two gauge plots (one for the red color pattern in  $B_{0,0}$  and other for the pattern in  $B_{1,1}$ ) (Figure 1(c)(iv)). These plots show the percentage of ICU patients who took sodium bicarbonate as input. From the two plots, the user can conclude that the number of ICU survivors who took sodium bicarbonate is less in number compared to non-survivors (effect of sodium bicarbonate on mortality rate is discussed in paper<sup>4</sup>).

### Conclusion

We demonstrated the usefulness of our system to detect and visualize the parameters which are more susceptible to the mortality. Though our results are set of rows, it is difficult for the end user to apriori decide on such complex SQL queries to get the results. Our visual interface helps in data exploration with implicit semantics as directed by the user.

### References

- [1] Vadapalli S, Karlapalem K. Heidi matrix: Nearest neighbor driven high dimensional data visualization. In Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration 2009 Jun 28 (pp. 83-92). ACM.
- [2] Gershon N, Eick SG. Visualization's new tack: making sense of information. IEEE spectrum. 1995 Nov;32(11):38-40.
- [3] Johnson AE, Pollard TJ, Shen L, Li-wei HL, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-III, a freely accessible critical care database. Scientific data. 2016 May 24;3:160035.
- [4] Kim HJ, Son YK, An WS. Effect of sodium bicarbonate administration on mortality in patients with lactic acidosis: a retrospective analysis. PloS one. 2013 Jun 5;8(6):e65283.