**MIMIC2V30**

In the existing version V2.6, we included the data for patients, who admitted our collaborating hospital between 2001 to 2008. In the updated V3.0, we expand the dataset to include data for additional patients, who were admitted between 2008 to 2012. That is, we have about 12 years of ICU patient data now. Not only more patients were included in the updated V3.0, more data variables were included as well. In V3.0, we added in the first three digit of Zip Codes as part of the demographic information of patients; pre-admission lab test results were also added to be used as baseline values for studies; physician notes were added as new components of the free text data; and more patient height data were included to enable a more accurate estimation of patients’ BMI.

In terms of data organization, we have consolidated all our nomenclatures in a single dictionary table. We are also working to clean up and map our nomenclatures to standardized coding systems, such as SNOMET CT, LOINC, RXNORM and CPT.

Another paradigm shift that we have introduced with the new release V3.0 is that we plan to harness the power of crowd sourcing for the continuous update of documentation and data. Github repository and some other discussion forums were explored to encourage users of MIMIC-II to actively participate and contribute to the continuous update of the dataset documentation and the dataset itself.

**EICU Dataset**

We have received around 1 TB of data from the Philips eICU Research Institute. The current data dump from Philips contains the data from around 2.5 million patients from over 800 ICUs in US. We have already imported the data into our database, and the following are the action plans for the next 6 months:

* Data exploration: generate basic statistics of the eICU data so to gain a rough understanding of the contents in the dataset
* Data quality assessment: develop data quality metrics and assessment rules so that we can evaluate and assign a quality score to the data from individual ICUs.
* Open data subset selection: based on the quality of the data, select 100K patients with the best data quality to be released to the research community

**Conferences and Demos by Mornin:**

* Conducted a system demo on “Big Data for Critical Care with Cloud-based In-memory Database” at AMIA 2014. The demo showcased how we manage our MIMIC-II dataset, which consists of patients’ clinical data, vital sign time series data and free text notes data, with cloud-based in-memory database management system (DBMS). We implemented the solution with SAP’s HANA in-memory DBMS and Amazon’s AWS cloud. Our demo was recognized as the daily highlight of AMIA 2014.
* Organized a tutorial on “Management and Analytic of Biomedical Big Data with Cloud-based In-Memory Database and Dynamic Querying” at the Knowledge Discovery and Data Mining (KDD) 2014 Conference (one of the best conferences in data mining). The tutorial was well received with over 20 participants. During the tutorial, we discussed the unique challenges in biomedical big data management and analytics. We also investigated on how we can move analytics close to the data to increase computational performance and promote better data security.