TECHNOLOGY BUCKET	:	**HEALTHCARE & BIOMEDICAL DEVICES**
COMPANY NAME	:	**ezDI**
TEAM NAME	:	**THE_OUTLIERS**
TEAM LEADER NAME	:	**APOORVA VIKRAM SINGH**
PROBLEM STATEMENT	:	**PATIENT CASE SIMILARITY**
PROBLEM CODE	:	**RG2**
CATEGORY	:	**SOFTWARE**
COLLEGE CODE	:	**U-0055**

THE SOLUTION:

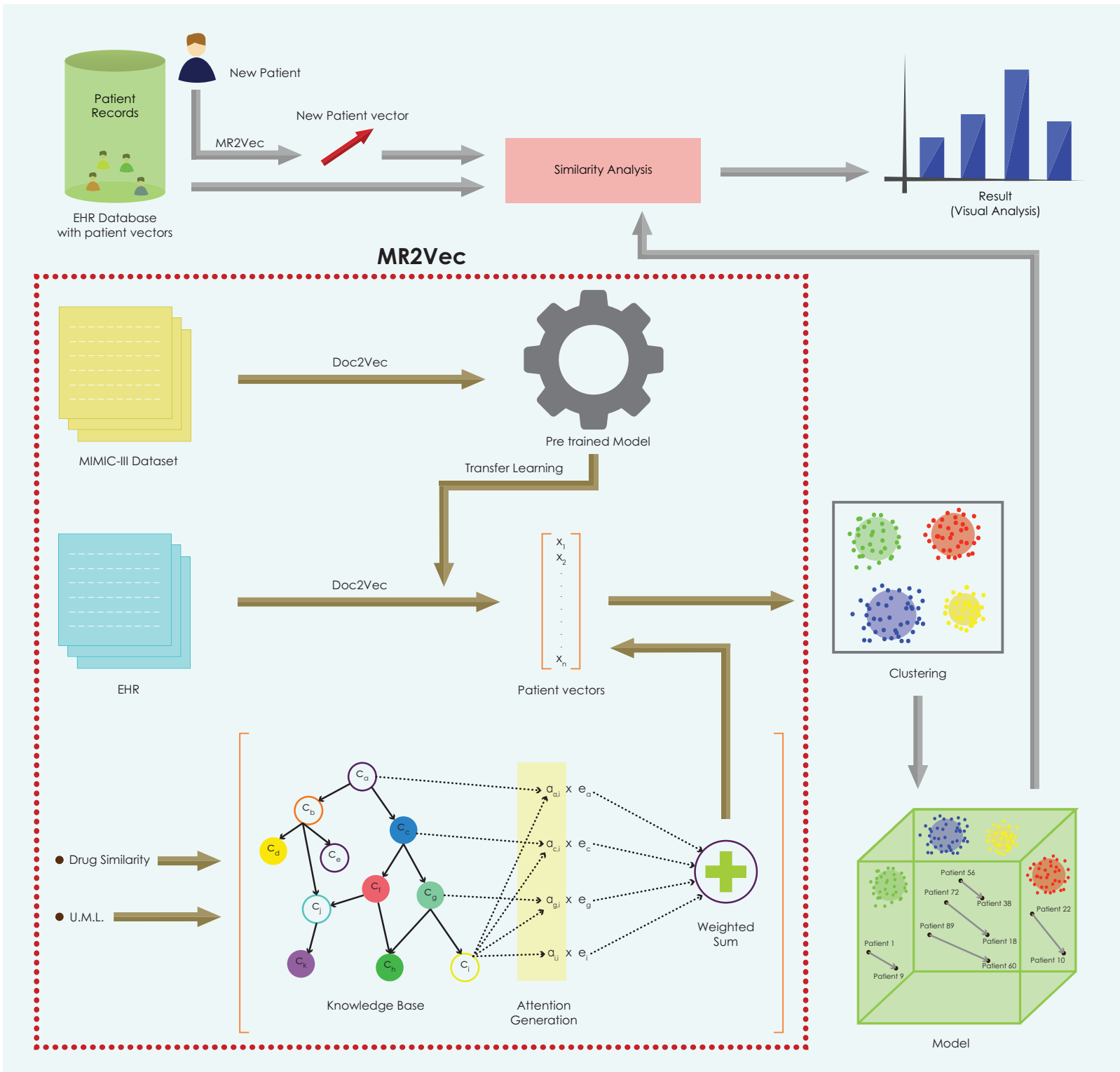
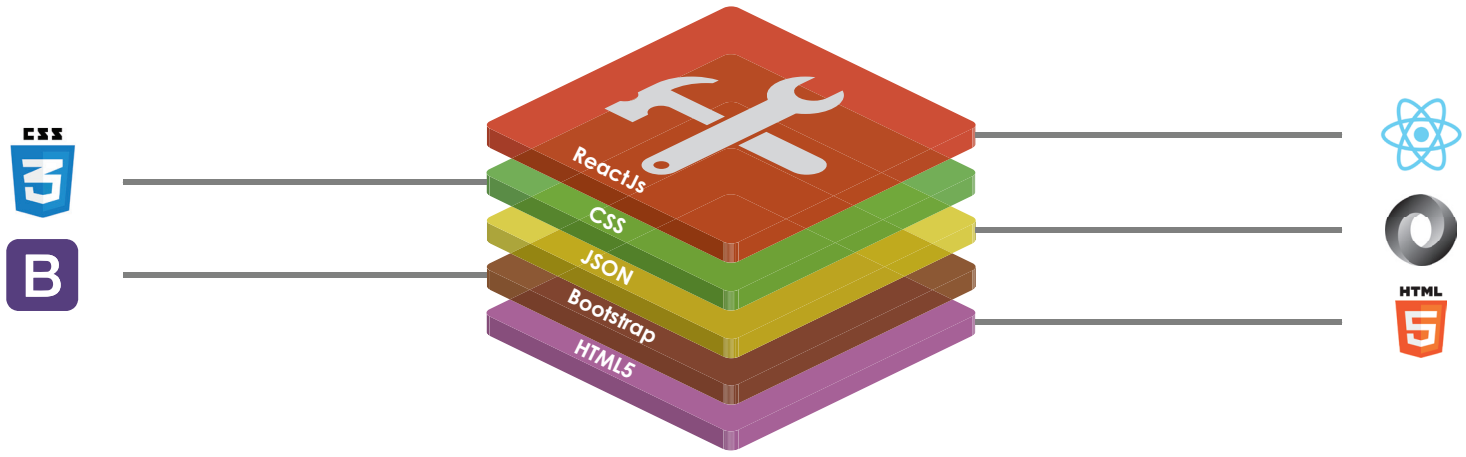


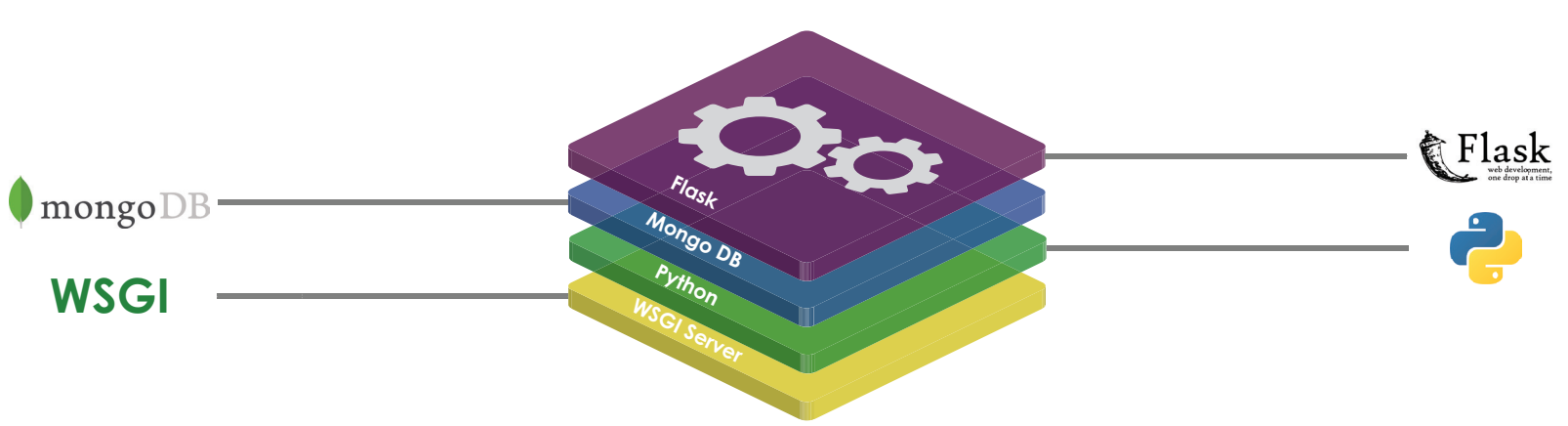
FIG. 1 : PIPELINE FOR THE MACHINE LEARNING MODEL.

- **MR2Vec :** A novel vectorization technique to create patient vectors.
- **MR2Vec** = Doc2Vec on (MIMIC-III + ezDI) dataset + Transfer Learning + Knowledge Base (UMLS database + DrugBank database) + Attention Model
- **Transfer Learning :** To re-train primary model (obtained by deploying Doc2Vec on MIMIC-III) on ezDI dataset to eliminate irrelevant noise.
- **Knowledge Base (KB) :** Assembled by weights obtained from UMLS data and Drug-Bank data, these weights are further multiplied with attention parameters to obtain weighted sums.
- **Clustering :** Generating patient vector clusters on basis of similarity index.
- **Triplet-Loss Metric :** To calculate normalized distance metric between patient vectors as a measure for similarity score.
- **Interactive Interface :** Leveraging this machine learning model to engine a website that eases patient similarity analysis practices for Medicare facilities.

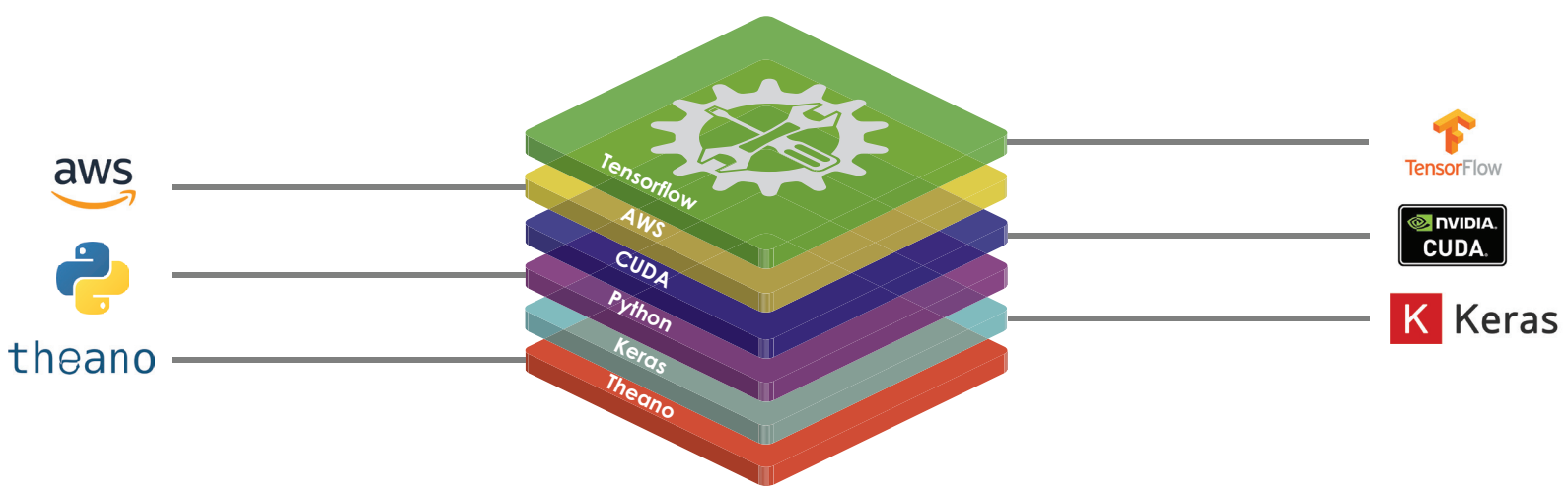
TECHNOLOGY STACK



Client-Side Programming

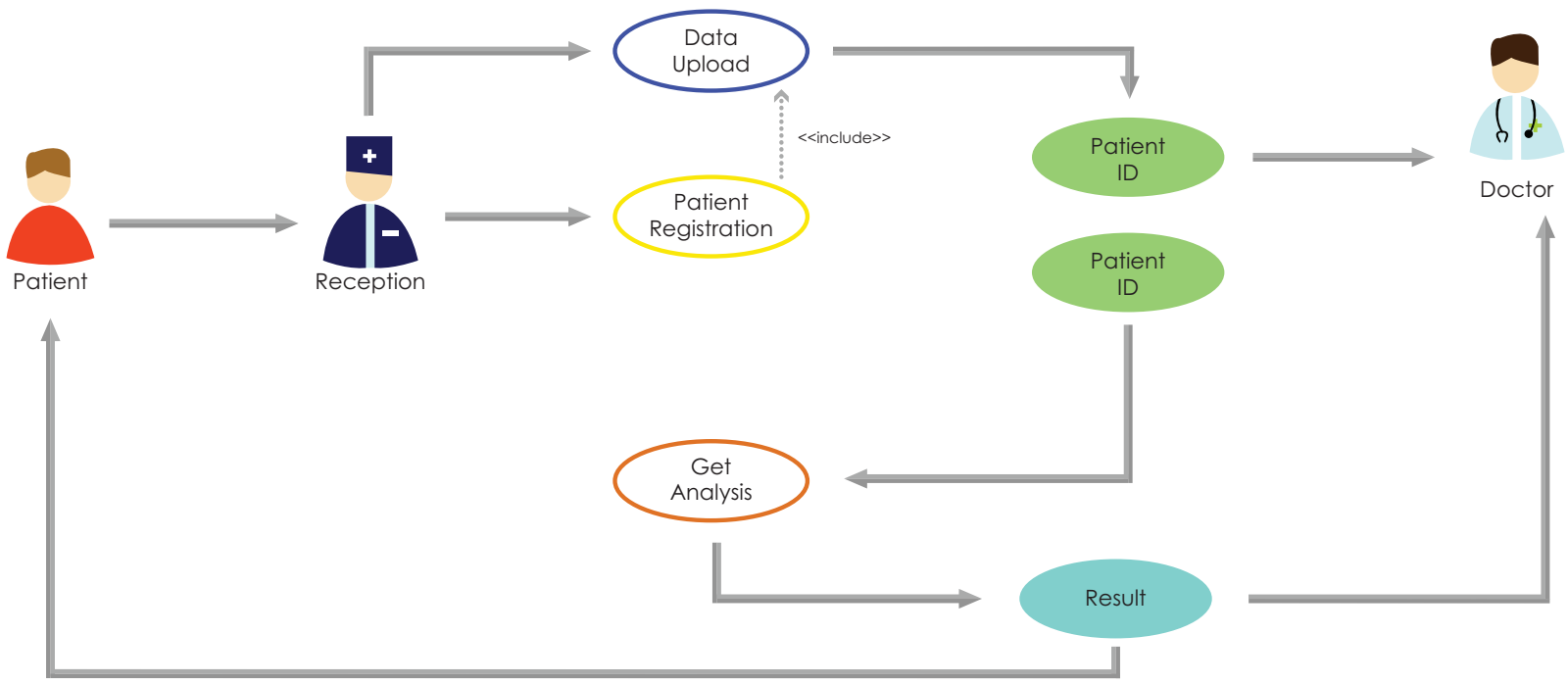


Server-Side Programming

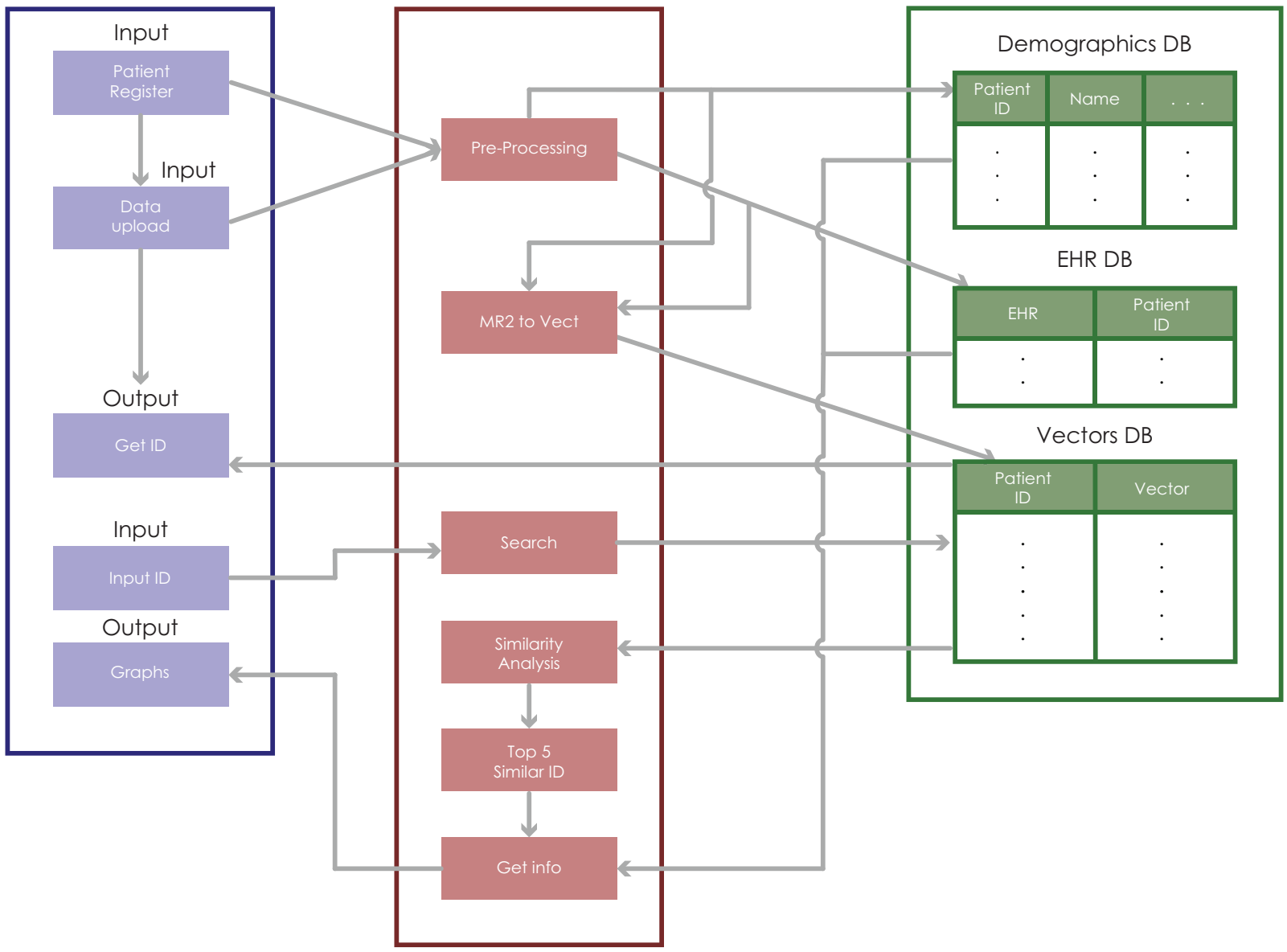


Machine Learning Tools

USE-CASE DIAGRAM



INTERFACE-STRUCTURE DIAGRAM



SHOW-STOPPER

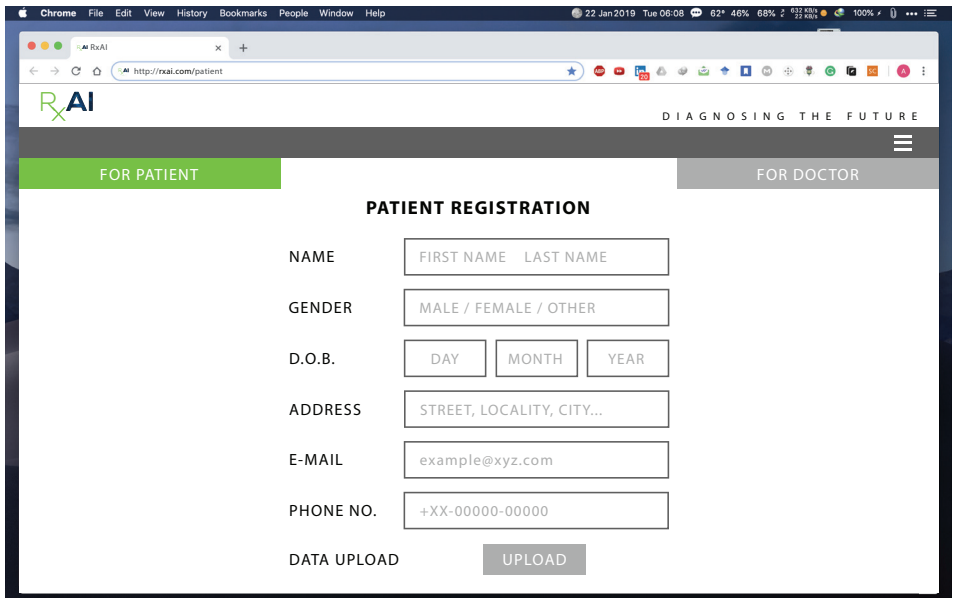


FIG. 2.1 : INTERFACE FOR DATA UPLOAD

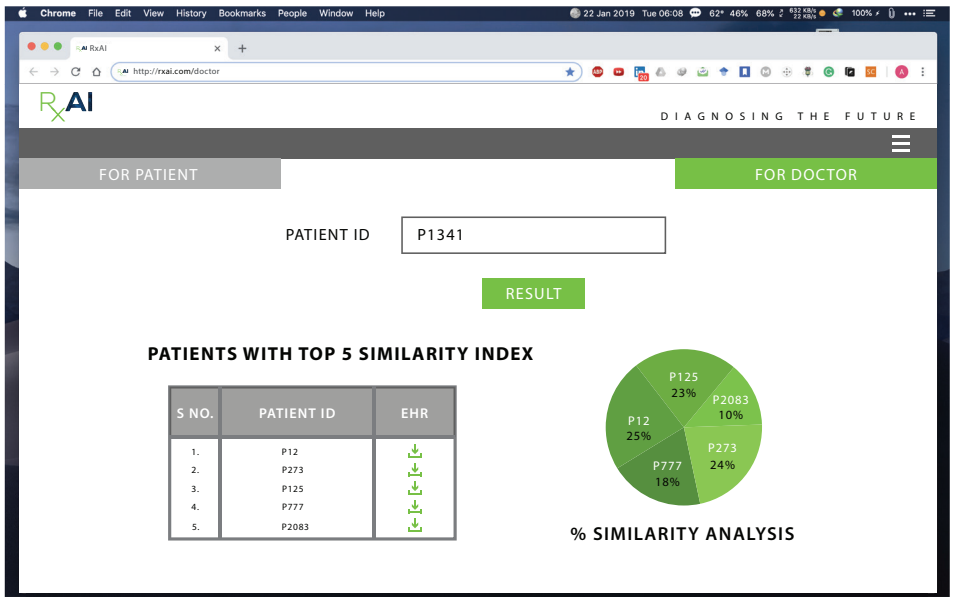


FIG. 2.2 : INTERFACE FOR SIMILARITY ANALYSIS(DOCTOR)

- The machine learning model used for predictive analysis of EHR data employs the novel MR2Vec technique for creating vectors of medical records of inpatients.
- MR2Vec utilizes Doc2Vec based intuitive scheme to generate a primary model. This primary model is leveraged to train on ezDI dataset to yield a secondary model which will be a fine-tuned version of primary model trained on relatively larger MIM-IC-III dataset.
- MR2Vec provides much-improved results when compared to generic Word2Vec model. It even provides an intuitive edge over Word2Vec model implemented with pre-trained embeddings which fails to capture thorough semantics from data.
- To further improve the embeddings, we have exploited UMLS database and Drug-Bank to assemble a Knowledge Base (KB) which will yield a weighted sum by multi-plying weights with attention parameters.
- We aim to establish an interactive client - server interface that proffers a compre-hensive anatomization of patient's chronic health and aids Medicare forums in diagnosing the inpatients.