

As the adoption of electronic health records (EHRs) has grown, EHRs are now composed of a diverse array of data, including structured information and unstructured clinical progress notes. Two unique challenges need to be addressed in order to utilize EHR data in clinical research and practice: 1) Computational phenotyping: How to turn complex and messy EHR data into meaningful clinical concepts or phenotypes? 2) Predictive modeling: How to develop accurate predictive models using longitudinal EHR data? To address these challenges, I will present our approaches using a case study on early detection for heart failure. For computational phenotyping, we present EHR data as data as inter-connected high-order relations i.e. tensors (e.g. tuples of patient-medication-diagnosis, patient-lab, and patient-symptoms), and then develop expert-guided sparse nonnegative tensor factorization for extracting multiple phenotype candidates from EHR data. Most of the phenotype candidates are considered clinically meaningful and with great predictive power. For predictive modeling, I will present how using deep learning to model temporal relations among events in EHR improved model performance in predicting heart failure (HF) diagnosis compared to conventional methods that ignore temporality.