

Appendix: Summary of Best Papers Selected for the 2018 Edition of the IMIA Yearbook, Section AI in Health

Albers DJ, Levine ME, Stuart A, Mamykina L, Gluckman B, Hripcsak G

Mechanistic machine learning: how data assimilation leverages physiological knowledge using bayesian inference to forecast the future, infer the present, and phenotype

J Am Med Inform Assoc 2018;25(10):1392-401

Albers *et al.*, applied the computational technique of data assimilation to predict glucose values, input missing data, and learn model parameters that represent phenotypes in type 2 diabetes. Data assimilation is an established technique that has been used to solve problems in space travel. In this medical application, researchers combined machine learning with a mechanistic model of endocrine physiology that constrained the space for learning and allowed accurate prediction of future blood glucose levels, based on finger-prick glucose measurements and data about meal carbohydrates. The approach performed well even in the context of sparse data, which is a common problem in health applications. The potential for data assimilation to identify physiologically meaningful phenotypes was demonstrated through estimation of distinct model parameters for diabetic and non-diabetic individuals. Incorporation of physiological domain knowledge with machine learning algorithms supports trust and explainability, which are important

for user adoption of AI solutions. The authors beautifully articulated the need for a multidisciplinary biomedical informatics approach and pipeline to leverage the data assimilation method in healthcare.

Oktay O, Ferrante E, Kamnitsas K, Heinrich M, Bai W, Caballero J, Cook SA, de Marvao A, Dawes T, O'Regan DP, Kainz B, Glocker B, Rueckert D

Anatomically Constrained Neural Networks (ACNNs): application to cardiac image enhancement and segmentation

IEEE Trans Med Imaging 2018;37(2):384-95

Oktay *et al.*, presented a novel strategy to train neural networks to learn shape models of underlying anatomy from cardiac imaging data using a stacked autoencoder constrained by global anatomic knowledge. These anatomically constrained neural networks demonstrated excellent performance in image segmentation of multimodal two-dimensional magnetic resonance imaging (MRI) and three-dimensional ultrasound cardiac imaging data sets (i.e., the United Kingdom Digital Heart Project Dataset, the Challenge on Endocardial Three-dimensional Ultrasound Segmentation (CETUS) 2014 Challenge Dataset, and the Automated Cardiac Diagnosis Challenge (ACDC) Medical Image Computing and Computer-Assisted Intervention Society (MICCAI) 2017 Challenge Dataset). In addition to segmentation, the algorithm showed good performance for the task of pathological classification. This approach addressed the common problems of motion artifact, data corruption, and a lack of internal consistency, which limit many approaches to image interpretation, and it could be generalized to other image processing tasks for which prior knowledge can provide guidance to learning models.

Lee J, Sun J, Wang F, Wang S, Jun CH, Jiang X

Privacy-preserving patient similarity learning in a federated environment: development and analysis

JMIR Med Inform 2018;6(2):e20

Lee *et al.*, addressed the important and challenging task of enabling learning in a federated environment in healthcare, where sufficient data to make inferences might be stored across a wide variety of institutional sites. The authors presented a novel privacy-preserving analytics platform for learning patient similarity in a federated setting, using a multi-hash approach for context-dependent, cross-institution patient representation with incorporation of homomorphic encryption for privacy preservation. The authors evaluated their approach using sequential medical events extracted from the Multiparameter Intelligent Monitoring in Intensive Care (MIM-IC)-III database, a collection of data from patients admitted to the intensive care unit at Beth Israel Deaconess Hospital, to predict the incidence of five diseases common in critically ill patients: disorders of lipid metabolism, hypertensive chronic kidney disease, cardiac dysrhythmias, heart failure, and acute renal failure. They compared the performance of their federated approach with an open system, where patient data can be shared without restrictions across sites, and a closed system, where data can only be used locally. Their system with a multi-hash approach exceeded closed and approximated open system performance, and calculated similarities across sites remained the same after homomorphic encryption, demonstrating preservation of privacy.