

Editorial: Predictive Intelligence in Biomedical and Health Informatics

BIG data is fueling diverse research directions in both medical image analysis and computer vision research fields. These can be divided into two main categories: (1) analytical methods, and (2) predictive methods. While analytical methods aim to efficiently analyze, represent, and interpret data, predictive methods leverage the data currently available to predict observations at present (e.g., by completing missing observations), at previous time-points (e.g., by solving reverse problems), or at later time-points (i.e., forecasting the future). For instance, a method which only focuses on classifying patients with mild cognitive impairment (MCI) and patients with Alzheimer’s disease (AD) is an analytical method, while a method that predicts if a subject diagnosed with MCI will remain stable or convert to AD over time is a predictive method. Similar cases can be established for various neurodegenerative or neuropsychiatric disorders, degenerative arthritis, or cancer studies, in which the disease/disorder develops over time.

It would constitute a stunning progress in the biomedical data analysis research community if, in a few years, we contribute to engineering a ‘predictive intelligence’ which can map both low-dimensional and high-dimensional medical data into the future with high precision. This special issue is the first endeavor to drive the field of ‘high-precision predictive medicine’, where late medical observations are predicted with high precision, while providing interpretable explanations via machine learning and deep learning based on statistical, mathematical or physical models of healthy, disordered development and aging subjects. Despite the notable progress of analytical methods within the last twenty years in areas such as medical image segmentation, registration or other related applications, efficient predictive intelligent models/methods are still lagging behind. As such, predictive intelligence presents novel solutions with the potential to surpass existing methods, and is more likely to increase exponentially over the next few years. This will have far-reaching consequences for the development of new treatment procedures and state-of-the-art technologies. These predictive models will begin to shed light on complex healthcare and medical challenges we have ever encountered, and, in doing so, change our basic understanding of who we are.

The main aim of this special issue is to propel the advent of predictive models in a broad sense, with application to medical data. Particularly, this issue will focus on papers describing new

cutting-edge predictive models and methods that solve challenging problems in the medical field. Topics of interest include predictive methods dedicated to modeling and predicting disease development or evolution from a limited number of observations; computer-aided prognostic methods (e.g., for brain diseases, prostate cancer, cervical cancer, dementia, acute disease, and neurodevelopmental disorders); forecasting disease/cancer progression over time; predicting low-dimensional data (e.g., behavioral scores, clinical outcome, age, gender); predicting the evolution or development of high-dimensional data (e.g., shapes, graphs, images, patches, abstract features, learned features, and deep data representations); predicting high-resolution data from low-resolution data; prediction methods using 2D, 2D + t, 3D, 3D + t, ND and ND + t data; predicting an image from a different modality image (e.g., data synthesis); predicting lesion evolution; predicting missing data (e.g., data imputation or data completion problems).

This special issue is based on the advances in machine learning and statistical modeling of predictive methods, applied to biomedical applications. Seventeen manuscripts were accepted for publication in this issue after several rounds of blind peer-review. In summary, each of these papers describes a method for predicting clinical scores, variables, or diagnosis based on the input biomedical data including magnetic resonance images (MRIs).

Kassani *et al.* presented a sparse machine learning solution across whole-brain functional connectivity (FC) measures of three data sets, derived from resting state functional magnetic resonance imaging (rsfMRI) and two tasks of fMRI data including a working memory nback task (nb-fMRI) and an emotion identification task (emfMRI).

Hui and Bath introduced a large and wide variety of factors from multiple domains using a large nationally representative sample of older people from the English Longitudinal Study of Ageing (ELSA). Seven machine learning algorithms were implemented to build predictive models for performance comparison. Individual-based model prediction results were later combined via a simple model ensemble for improved performance. A series of important factors in each domain area were identified drawing attention to new evidence of a likely association with the onset of dementia later in life. This information will advance our understanding of potential risk factors for dementia and identify warning signs of the early stages of dementia.

Teng *et al.* evaluated the prognostic value of lymph node ratio (LNR) for the survival of breast cancer patients using Bayesian

inference. A prognostic modeling framework was proposed using Bayesian inference to estimate the impact of LNR in breast cancer survival. Based on the proposed model, they developed a web application for estimating LNR and predicting overall survival. The final survival model with LNR outperformed the other models considered (C-statistic 0.71). Compared to the directly measured LNR, the estimated LNR slightly increased the accuracy of the prognostic model. Model diagnostics and predictive performance confirmed the effectiveness of Bayesian modeling and the prognostic value of the LNR in predicting breast cancer survival. The estimated LNR was found to have a significant predictive value for the overall survival of breast cancer patients. Specifically, they used Bayesian inference to estimate LNR which was then used to predict overall survival.

Casamitjana *et al.* used a multivariate approach based on Projection to Latent Structures in its Regression variant (PLSR) to study structural changes related to aging and AD pathology. MRI volumetric and cortical thickness measurements were used for brain morphology, and cerebrospinal fluid (CSF) biomarkers (t-tau, p-tau and amyloidbeta) were used as a proxy for characterizing AD pathology. By relating both sets of measurements, PLSR found a low-dimensional latent space describing AD pathological effects on brain structure. The proposed framework allowed to separately model aging effects on brain morphology as a confounder variable orthogonal to the pathological effect.

Xue *et al.* proposed an MRI super-resolution (SR) model named progressive sub-band residual learning SR network (PSR-SRN). The proposed model contains two parallel progressive learning streams, where one stream learns on missed high-frequency residuals by sub-band residual learning unit (ISRL) and the other focuses on reconstructing refined MR image. These two streams complement each other and enable to learn complex mappings between “Low-” and “High-” resolution MR images. Besides, they introduced brain-like mechanisms (in-depth supervision and local feedback mechanism) and progressive sub-band learning strategy to emphasize variant textures of MRI.

Hamad *et al.* proposed a new data-driven approach that aims to increase precision and sensitivity in human activity recognition applied in a smart home setting applied to healthcare applications. The proposed method considered the partial oncoming sensor activations in addition to preceding sensor activations. With the use of oncoming sensor activation, they could take the benefits of enhancing the learning process that leads to improved recognition performance compared with the approaches using only the preceding sensor activations in the intelligent environment.

Madan *et al.* argued that latent biomarkers are quantities that strongly relate to patient’s disease diagnosis and prognosis but are difficult to measure or even not directly observable. Hence, they developed, analysed, and validated new priors for Bayesian inference of such biomarkers. Theoretical analysis revealed a relationship among the estimates inferred from the model, the true values of measured quantities, and the impact of the priors.

Jalali and Lee discussed the importance atrial fibrillation (AF) as the most prevalent cardiac arrhythmia. The atrial beat is irregular during AF, which hardly causes any blood flow. This may cause blood clot formation and cardioembolic strokes. Thus,

computer-aided devices may assist cardiologists in diagnosing heart rhythm disorders better. They attempted to identify the premature atrial complexes (PACs) to predict the occurrence of AF by using electrocardiogram (ECG) spectrograms.

Li *et al.* argued that, for people with Type 1 diabetes (T1D), forecasting blood glucose (BG) can be used to effectively avoid hyperglycemia, hypoglycemia and associated complications. The latest continuous glucose monitoring (CGM) technology allows people to observe glucose in real-time. They introduced GluNet, a framework that leverages a personalized deep neural network to predict the probabilistic distribution of short-term (30-60 minutes) future CGM measurements for subjects with T1D based on their historical data including glucose measurements, meal information, insulin doses, and other factors.

Jarrett *et al.* developed a novel convolutional approach that addresses the drawbacks of both traditional and statistical approaches as well as recent neural network models for survival analysis. Specifically, they presented MATCH-Net: a Missingness-Aware Temporal Convolutional Hitting-time Network, designed to capture temporal dependencies and heterogeneous interactions in covariate trajectories and patterns of missingness.

Shamout *et al.* proposed the ‘Deep Early Warning System’ (DEWS), an interpretable end-to-end deep learning model that interpolates temporal data and predicts the probability of an adverse event, defined as the composite outcome of cardiac arrest, mortality or unplanned ICU admission. The model was developed and validated using routinely collected vital signs of patients admitted to the Oxford University Hospitals between 21st March 2014 and 31st March 2018.

Yu and Xie argued that hospital readmission is one of the most critical issues in the healthcare system due to its high prevalence and cost. The improvement effort necessitates reliable prediction models which can identify high-risk patients effectively and enable healthcare practitioners to take a strategic approach. Using predictive analytics based on electronic health record (EHR) for hospital readmission is faced with multiple challenges such as high dimensionality and event sparsity of medical codes and the class imbalance. They proposed a data-driven analytical framework using hospital inpatient administrative data from a nationwide healthcare dataset.

Guo *et al.* argued that Low-Dose CT (LDCT) can significantly improve the accuracy of lung cancer diagnosis and thus reduce cancer deaths compared to chest X-ray. The lung cancer risk population is also at high risk of other deadly diseases, for instance, cardiovascular diseases. Therefore, predicting the all-cause mortality risks of this population is of great importance. They introduced a knowledge-based analytical method using deep convolutional neural network (CNN) for all-cause mortality prediction. The underlying approach combines structural image features extracted from CNNs, based on LDCT volume at different scales, and clinical knowledge obtained from quantitative measurements, to predict the mortality risk of lung cancer screening subjects.

Zhang *et al.* proposed methods for epilepsy seizure prediction that paves the way of timely warning for patients to take more active and effective intervention measures. Compared to

seizure detection that only identified the inter-ictal state and the ictal state, fewer works were conducted on seizure prediction. They proposed a novel solution on seizure prediction using common spatial pattern (CSP) and convolutional neural network (CNN).

Wang *et al.* proposed a novel deep CNN called Thorax-Net to diagnose 14 thorax diseases using chest radiography. Thorax-Net consists of a classification branch and an attention branch. The classification branch serves as a uniform feature extraction-classification network to free users from the tedious hand-crafted feature extraction and classifier design. The attention branch exploits the correlation between class labels and the locations of pathological abnormalities via the analysis of the feature maps learned by the classification branch.

Yu *et al.* argued that estimating hospital mortality of patients is important in assisting clinicians to make decisions and hospital providers to allocate resources. They proposed a multi-task recurrent neural network with attention mechanisms to predict patients' hospital mortality, using reconstruction of patient physiological time series as an auxiliary task.

Finally, He *et al.* used Multispectral Imaging (MSI) that provides a sequence of en-face fundus spectral slices and allows for the examination of structures and signatures throughout the thickness of retina to characterize diabetic retinopathy (DR) lesions comprehensively. Manual interpretation of MSI images is commonly conducted by qualitatively analyzing both the spatial and spectral properties of multiple spectral slices. Meanwhile, there were a few algorithms that can effectively exploit the

spatial and spectral information of MSI images for the diagnosis of DR. They proposed a new approach that can quantify the spatial spectral features of MSI retinal images for automatic DR lesion segmentation.

The guest editors wish to thank all the authors for sharing their research and the reviewers for helping to improve the works published here. They hope that novel approaches as represented by accepted articles in this special issue introduce new cutting-edge predictive models and methods that solve challenging biomedical problems.

E. ADELI, *Guest Editor*
Stanford University
Stanford, CA 94305 USA

I. REKIK, *Guest Editor*
Istanbul Technical University
34467 Istanbul, Turkey

S. H. PARK, *Guest Editor*
Daegu Gyeongbuk Institute of
Science and Technology
Daegu 42988, South Korea

D. SHEN, *Guest Editor*
The University of North
Carolina at Chapel Hill
Chapel Hill, NC 27599 USA