Diagnosis of acute bacterial and viral infections using a 29-mRNA neural network

Michael Mayhew, PhD[1], Evangelos J. Giamarellos-Bourboulis, MD[2],, Ljubomir Buturovic, PhD[1], Roland Luethy, PhD[1], Damjan Krstajic, MSc[1], Kirindi Choi, BA[1], Purvesh Khatri, PhD[1], Oliver Liesenfeld, MD[1], Charalambos Gogos, MD[3], Timothy E. Sweeney, MD, PhD[1]

[1]Inflammatix, Inc. Burlingame, CA, USA; [2]4th Department of Internal Medicine, National and Kapodistrian University of Athens, Athens, Greece; [3]Department of Internal Medicine, University of Patras, Greece

Background

Diagnosis of acute bacterial and viral infections remains a difficult clinical challenge. Biomarker-based testing can be more sensitive than pathogen detection, as the immune response represented in peripheral blood leukocytes can sense infections anywhere in the body. However, multi-biomarker panels have traditionally suffered from overfit, non-robust statistical models. Here, we present a 29-mRNA neural network ('HostDx Sepsis') trained to diagnose bacterial and viral infections across multiple transcriptomic datasets and then applied with fixed weights across platform transfer into independent data.

Methods

We pre-processed training data using open-source COCONUT software to reduce technical heterogeneity among different sample sources and gene expression platforms. Using 18 normalized training datasets from public sources (N=1,092), we developed a three-class neural network classifier which assigns an input sample to a bacterial, viral or non-infected category, using 29-gene expression profile as input and reporting the class probabilities as output. We also evaluated other state-of-the-art machine learning classifiers, but none achieved clinically-viable performance on the training data. We then applied the neural network into a prospective cohort in Greece (PROMPT/NCT 03295825) of ED patients with suspected infections, profiled on NanoString nCounter.

Results

In leave-one-dataset-out cross-validation, our 29-mRNA neural network achieved a bacterial-vs-other (BAO) AUROC of 0.91 and a viral-vs-other (VIO) AUROC of 0.90, vs. 0.89/0.84 for the best Support Vector Machine classifier. We then applied the final fixed-weight model trained on all 18 datasets into a new prospective cohort (PROMPT) with bacterial or viral infections or non-infectious illness. In PROMPT the BAO AUROC was 0.91, and the VIO AUROC was 0.87. Cutoffs set in the LODO pooled data for target likelihood ratios (LRs) of 0.06 (rule-out) and 10.1(rule-in) yielded similar LRs in PROMPT (lower LR 0, upper LR 9.6).

Conclusions

HostDx Sepsis has demonstrated clinically useful performance for classification of acute infections. To our knowledge, this is the first report of a neural network significantly outperforming standard machine learning methods on a challenging and clinically relevant genomic classification problem, and maintaining the clinically strong performance on an independent test data without any retraining.

The study was funded in part by the Hellenic Sepsis Study Group