

From Dream to Diagnosis: How BI Delivered A High-Accuracy Predictive Model For Cancer Detection



SITUATION

- A government-funded cancer center sought to implement an Artificial Intelligence (AI)-powered classifier, utilizing deep neural networks (DNN), for diagnosing cancer using RNA-sequencing data (RNA-seq).
- The Cancer Center had a gap lacked the specific in-house resources to develop such a complex predictive model, so they turned to a trusted partner for help.

STRATEGY & PROCESS

- Bridge Informatics (BI), a leading bioinformatics data management and analysis company, partnered with the Cancer Center to help bring their AI dreams to life.
- BI leveraged its expertise in data mining and deep learning for cancer classification to create a Statement of Work that included the following:
 - **Data Acquisition:** BI accessed cancer-specific bulk transcriptional data from The Cancer Genome Atlas (TCGA) database. We then implemented a filtering process to ensure sufficient patient samples (> 150 patients), with tumor-matched normal tissue, were present for the cancers of interest to train and test the DNN. This resulted in a refined dataset encompassing nine human cancers.
 - **Feature Selection:** BI performed differential gene expression analysis using a trusted algorithm to identify genes exhibiting significant differences between tumor and tumor-matched normal control samples. The resulting gene lists were filtered after carefully choosing a relevant cutoff threshold and acceptable false discovery rate with an appropriately adjusted p-value.
 - **Deep Neural Network Development:** The resulting dataset, consisting of differentially expressed genes as inputs and cancers as labels, was used for training and testing the DNN architecture. The network architecture consisted of 339 gene inputs, representing the genes passing cutoffs, followed by three hidden layers and proprietary activation functions for complex information processing. Finally, the network culminated in 10 distinct outputs, each corresponding to one of the nine cancer types identified earlier, with an additional output for normal tissue classification.

RESULTS

- Within six months, BI equipped the Cancer Center with a powerful deep neural network AI cancer classifier that was trained on the TCGA's real patient data.
- The AI architecture was refined to distinguish cancer from normal tissue with >90% and an F1 score of > 0.8.
- The deep neural network, trained on the curated RNA-seq data, is designed to analyze patient samples and predict cancer presence and type with high accuracy, empowering oncologists with valuable insights to support their diagnoses and potentially improve patient outcomes.