The Guided Inquiry Neural Network (GINN): Developing A Novel Machine Learning Architecture for Differential Diagnosis in Primary Care Settings Utilizing Data Masking, Bayesian Inference, & Feature Importance

Jain, Saveer (School: duPont Manual High School)

Primary care physicians must determine which questions to ask and which tests to run to arrive at their diagnosis. Current AI architectures are not suited for such diagnoses because they (1) require a pre-determined combination of inputs, and (2) produce results that are uninterpretable. This project developed a novel architecture, the Guided Inquiry Neural Network (GINN), to ask a patient a line of relevant questions and generate a differential diagnosis. The GINN uses feature importance and posterior probabilities to drive an “information gain” function for question selection and was trained on a large, symptom-to-disease medical dataset that was masked to simulate incomplete knowledge about patients. Model performance was evaluated on question efficiency and accuracy per information gained. Questions were found as efficient: in just its first 10 questions, the GINN yielded an average of 6.63 affirmed questions and already retrieved 65.5% of the patients’ symptom data. Moreover, the model proved accurate in diagnosing the patient with this limited information; for samples with under 50% of known patient symptoms, the GINN’s average precision, recall, and F1 score for 49 diseases exceeded that of a baseline neural network by over 20% and continued to exceed the baseline with more information until approximately evening out at complete information with an average F1 of 0.922. Future research should focus on dataset creation and incorporating negative data (“lack” of symptom) and symptom correlations into question evaluation. Paired with a physician, GINNs could broaden access to primary care and improve reliability and cost efficiencies in diagnoses.

Awards Won: