AI-Driven Analysis of Diagnostic Profiles in COVID-19 Patients: Implications for Healthcare Interventions

HIROKI MATSUO[1], ANDREAS STAMATIS[2,3], CHELSEA YAGER[4], ALI BOOLANI[5], & GRANT B. MORGAN[1]

[1]Education Psychology; Baylor University; Waco, TX; [2]Health & Sport Sciences; University of Louisville; Louisville, KY; [3]Sports Medicine; University of Louisville Health; Louisville, KY; [4]Neurology; Northwell Health Systems; Sleepy Hollow, NY; [5]Honors Department; Clackson University; Potsdam, NY

Category: Doctoral

Advisor / Mentor: Morgan, Grant (grant_morgan@baylor.edu)

ABSTRACT

The COVID-19 crisis has strained global healthcare systems, highlighting the significance of investigating comorbidities and secondary diagnoses in patients. Harnessing of data-driven insights, as facilitated by *artificial intelligence* (AI), has shown remarkable promise in enhancing the efficacy of healthcare strategies and ameliorating patient outcomes. **PURPOSE**: To identify diagnostic profiles in COVID-19 patients via AI-driven analysis, focusing on comorbidities and secondary diagnoses. **METHODS**: The analytical groundwork was established upon the scrutiny of 42,974 patients with PCR-confirmed COVID-19 diagnosis. Each record was characterized by 850 diagnostic indicators encompassing a spectrum of ailments, such as demyelinated diseases, seizure disorders, and various additional comorbidities. The predominant racial composition of the sample was White (n = 31,329, 73%). A majority of patients were of the female gender (n = 23,534, 55%). Data were collected using Electronic Medical Records through the Cerner system from 31 hospitals in a large health system. Finite mixture modeling, a form of model-based unsupervised machine learning, was employed to ascertain the presence of latent, distinguishable patterns among secondary diagnoses. Of the approximately 850 secondary diagnoses considered, 221 exhibited prevalence in over 50 patients. A sequence of mixture models was estimated, incrementally augmenting the number of latent profiles via maximum likelihood estimation with robust standard errors. Model solutions were subjected to rigorous evaluation, culminating in the selection of three diagnostic profiles predicated on statistical model-data fit, parsimony, and interpretability. **RESULTS**: The selected model revealed the presence of three distinct diagnostic profiles. These profiles were characterized by patients who: (1) exhibited a notably low likelihood of presenting with secondary diagnoses, (2) demonstrated heightened probabilities of manifesting commonly observed diagnoses within the United States, such as hypertension, hyperlipidemia, and a history of tobacco use, or (3) displayed elevated probabilities of harboring multiple comorbid diagnoses, spanning domains such as lung, heart, and kidney-related conditions. The initial profile encompassed 27,002 patients (63%), followed by the second profile comprising 11,419 patients (27%), and the third profile, accounting for 4,553 patients (11%). Patients were individually assigned probabilities denoting their affiliation with each profile, with respective average classification probabilities of .98, .89, and .94, signifying a high degree of classification confidence. **CONCLUSION**: Our findings demonstrate the potential application of AI in informing healthcare interventions, such as tailored treatment plans, early intervention, resource allocation, patient education, research and development, and healthcare policy.