Regeneron WESEF 2023 Finalist

Mia Hoxhaj

Medicine and Health Sciences

Predicting Patients at Risk for Preventable Adverse Events Upon Discharge from the Hospital with an Acute Exacerbation of Congestive Heart Failure

Congestive Heart Failure (CHF) is a chronic, progressive condition in which the heart muscle becomes enlarged, thick, and stiff, resulting in the inability to adequately pump from and/or fill its chambers. CHF is known to cause elevated rates of morbidity and mortality, acute resource utilization, and hospital readmission. Using a preexisting CHF patient database, descriptive and predictive statistical analyses were run through R software to identify the social and clinical factors which are most closely correlated with high rates of acute care resource utilization and CHF-related hospital readmissions. Three predictive models were then developed and compared to discover which analysis (K Nearest Neighbor, Regression Tree, or Random Forest) produced the most accurate model in predicting if patients were readmitted to the hospital. The factors within the database were collected based on the creation of a directed acyclic graph (DAG), which outlined all measurable contributing factors to CHF-related readmissions. Descriptive statistical analysis included a series of histograms and box plots, which identified patients who are African American, aged 80 to 95, and patients with a LACE+ readmission index score of 60 to 80, as individuals who are most likely to frequently utilize hospital resources. Initial analysis was run using an unsupervised machine learning algorithm, using techniques such as elbowing and clustering, followed by using supervised analysis through the 3 above predictive models. All analyses validated length of stay and total prior visits as influential features in the health status of CHF patients. Then, the 3 predictive algorithms were run to develop predictive models, and their accuracy was measured in terms of a calculation, a confusion matrix, and a root mean square (RMS) value. Variables were also classified by importance, and it was noted that LACE+ score, length of stay during the initial visit, and total prior visits were influential in a patient's readmission. The random forest model had the lowest RMS, identifying that model as the most accurate in making predictions for a patient's disposition. Ultimately, the various statistical analyses performed in the study indicate a basis for the development of a predictive analytics model to track patients' demographics and clinical history, predicting and preventing readmission by initiating timely interventions for at-risk patients. With these developments, healthcare systems can bridge the gaps in patient care created by social determinants of health, and focus on allocating resources towards the most vulnerable patient populations.