Utilizing machine learning to understand sepsis

Erik Hartman

Sepsis is one of the deadliest syndromes in modern time, with little to no effective therapies available. In this project, machine learning was utilized to gain insight into the underlying biology of sepsis - a necessary step in finding novel and effective treatments and diagnostic tools.

Sepsis is a syndrome (collection of symptoms) which is responsible for $\sim$ 20% of global deaths each year. It is extremely diverse and complex, rendering it difficult to both diagnose and treat. Recently, researchers have classified various types of sepsis and recognized that unique therapies are required for the different types. However, before creating treatments and diagnostic tools, we need to understand the underlying biology of the various types - which is easier said than done.

In this project, machine learning was utilized to understand the biology of a specific type of sepsis which is characterized by damage to the kidney (referred to as septic acute kidney injury or AKI). Machine learning is a way to make a machine find patterns in complex data - and sepsis as a disease can be seen as a complex mixture of biological molecules. Finding the pattern in this soup of molecules is therefore a task fit for machine learning, and could help us understand the disease.

An algorithm named a *biologically informed neural network* - that is: a machine learning algorithm (specifically a neural network) that reflects the underlying biology of the disease (hence biologically informed) was devised. Creating such an algorithm solves the black box problem in machine learning, which states that it is impossible to understand what a machine learning algorithm is doing when it is solving a problem. It also allows for introspection into the algorithm and understand what parts of the biology it finds important and interesting when analyzing specific types of sepsis.

The algorithm allowed for the finding of proteins and biological pathways which were important in classification of septic acute injury of different severity. It was generalized to be compatible with any type of condition, disease or syndrome. Further, the algorithm is available in a public repository, and anyone can now create a biologically informed neural network with one line of code. Therefore, it can be utilized in other experiments to not only progress in the development of treatments and diagnostics of sepsis - but also other diseases.