ABSTRACT

Rheumatoid Arthritis (RA) is a joint inflammatory autoimmune disease affecting 1% globally. Although not lethal, ~40% of RA patients are subjected to systemic manifestations and clinical complications of various involvements. Being without a cure, the ability to achieve remission of currently available treatments are dependent on immediate intervention. However, the complex nature of RA makes detection a highly personalized and time-consuming process. Most attempts to unravel the genetic complexities of RA have adopted the genome wide association studies (GWAS) method. However, critics have questioned GWAS' ability to identify true causal genes that aren't carried by associations to correlated variants due to linkage disequilibrium.

This study proposes a machine learning (ML) approach to identify a small subset of polymorphisms that can discriminate between RA patients and population control. 13 SNPs were identified to show remarkable predictive performances evident by the ability to achieve a consistent >0.9 on all performance metrics upon prediction using a 5-fold cross validation and 3 unseen test sets. This method was able to identify SNPs that were not previously found in associated to RA with various implications of functionality that can be explored.

Keywords: *Rheumatoid Arthritis, Machine Learning, Supervised Learning, SNPs, Prediction.*