

PREDICTING LYMPHOMA DEVELOPMENT BY EXPLOITING GENETIC VARIANTS AND CLINICAL FINDINGS IN A MACHINE LEARNING-BASED METHODOLOGY WITH ENSEMBLE CLASSIFIERS IN A COHORT OF SJÖGREN'S SYNDROME PATIENTS

ABSTRACT

Lymphoma development constitutes one of the most serious clinico-pathological manifestations of patients with Sjögren's Syndrome (SS). Over the last decades the risk for lymphoma genesis in SS patients has been studied aiming to identify novel biomarkers and risk factors predicting lymphoma development in this patient population. This project propose to genetic susceptibility profiles of SS patients along with known clinical, serological and histological risk factors enhance the accuracy of predicting lymphoma development in this patient population. The potential predicting role of both genetic variants, clinical and laboratory risk factors were investigated through a Machine Learning based (ML) framework which encapsulates ensemble classifiers. Ensemble methods empower the classification accuracy with approaches which are sensitive to minor perturbations in the training phase. The evaluation of the proposed methodology based on a 10-fold stratified cross validation procedure yielded considerable results in terms of balanced accuracy (GB: 0.7780 ± 0.1514 , RF Gini: 0.7626 ± 0.1787 , RF Entropy: 0.7590 ± 0.1837). *Conclusions:* The initial clinical, serological, histological and genetic findings at an early diagnosis have been exploited in an attempt to establish predictive tools in clinical practice and further enhance our understanding towards lymphoma development in SS.