Using Machine-Learning Methods for Allergen Amino Acid Sequence's Permutations

Authors : Kuei-Ling Sun, Emily Chia-Yu Su

Abstract : Allergy is a hypersensitive overreaction of the immune system to environmental stimuli, and a major health problem. These overreactions include rashes, sneezing, fever, food allergies, anaphylaxis, asthmatic, shock, or other abnormal conditions. Allergies can be caused by food, insect stings, pollen, animal wool, and other allergens. Their development of allergies is due to both genetic and environmental factors. Allergies involve immunoglobulin E antibodies, a part of the body's immune system. Immunoglobulin E antibodies will bind to an allergen and then transfer to a receptor on mast cells or basophils triggering the release of inflammatory chemicals such as histamine. Based on the increasingly serious problem of environmental change, changes in lifestyle, air pollution problem, and other factors, in this study, we both collect allergens and non-allergens from several databases and use several machine learning methods for classification, including logistic regression (LR), stepwise regression, decision tree (DT) and neural networks (NN) to do the model comparison and determine the permutations of allergen amino acid's sequence.

Keywords : allergy, classification, decision tree, logistic regression, machine learning
Conference Title : ICHIMBDS 2017 : International Conference on Health Informatics Management and Big Data Systems
Conference Location : Tokyo, Japan
Conference Dates : May 28-29, 2017