P-322

Variable selection for early diagnosis of congenital heart disease using random forest entropy calculations

Heath A., Gonzales M., von Alvensleben I.
Kardiozentrum, La Paz, Bolivia

Introduction: Decision trees have been widely used in order to measure the importance of variables. Nevertheless, in the field of bioinformatics their use is not widespread. Populations with high infant mortality need innovative strategies to recognize children at risk.
Methods: Variables present in medical records of patients with congenital heart disease (CHD) have been analysed, searching the most representative signs and symptoms which can lead physicians without experience in paediatric cardiology to recognize newborn and children with congenital heart disease and send them to a reference centre. Combining machine learning techniques and medical knowledge, twenty-eight variables assembled in a formulary based in answers yes/no were chosen as the most representatives for early recognition and diagnosis of cardiac malformation. By the combination of entropy and gain of information as parameters of decision trees it was possible to place variables in a structure according to their importance. The most recurrent signs and symptoms were repeatedly placed as root node of the different trees created for this research, getting as a result very high scores of accuracy in prediction. The use of these variables was backed up by a Pearson correlation matrix, whose results demonstrated the mathematic correlation between all of them.
Results: Entropy values for the variables with the highest correlation with CHD were: Down syndrome (0,99), recurrent pneumonia (0,99), clubbing (0,99), heart murmur (0,98), recurrent low oxygen saturation (0,97) and tachycardia (0,97). Variables with values lower than 0,7 do not help to predict congenital heart disease (e.g. splenomegaly).
Conclusions: This study may provide general practitioners and baseline paediatricians with a validated chart for the early suspicion of CHD and referral to opportune diagnosis and treatment.