Accurate stratification of patients with neuroblastoma by application of the reject option paradigm to outcome prediction

Davide Cangelosi (¹§), Martina Morini (¹), Alessandra Eva (¹), Maria Carla Bosco (¹), Massimo Conte (²) and Luigi Varesio (¹).

(¹) Laboratory of Molecular Biology, G. Gaslini Institute, Largo Gaslini 5, 16147 Genoa, Italy.
(²) Department of Hematology–Oncology, G. Gaslini Institute, Largo Gaslini 5, Genoa 16147, Italy.

§Corresponding author.

Present address: Laboratory of Molecular Biology, G. Gaslini Institute, Largo Gaslini 5, 16147 Genoa, Italy.

Phone: 0039 010 5636633

Email: DavideCangelosi@gaslini.org

ABSTRACT

Motivation

Neuroblastoma (NB) is the most common solid pediatric tumor, deriving from ganglionic lineage precursors of the sympathetic nervous system. Half of the patients with high-risk NB die despite treatment and more accurate outcome prediction models are needed to direct effective therapies. To limit the therapy's side effects, a highly accurate classification of a limited number of patients is preferable to a broader classification with greater error. We aimed at improving the classifier's accuracy...
and establishing the criteria for confident selection of classification rules by applying the reject option technique (RO).

**Methods**

We evaluated outcome prediction by BFTree, ID3, J48, REPTree, SimpleCART decision-tree algorithms utilizing a 182 patients’ dataset. Informed consent was obtained in accordance with institutional policies in use in each country. Fifty percent of the patients were utilized for model selection, and the remaining 50% for independent validation. The risk factors were: NB-hypo, age at diagnosis, stage and MYCN amplification. Accuracy of the classification was measured by Matthew’s Correlation Coefficient and assessed by 2 fold cross validation analysis repeated 1000 times. Trade-off between confidence and accuracy was estimated by the RO technique utilizing the accuracy/rejection plot. Kaplan-Meier estimate and log-rank test assessed overall survival.

**Results**

Every decision tree classified the patients' outcome with a confidence >0.6. Only ID3, utilizing all risk factors, stratified stage 4 patients and was chosen for further analysis. Application of RO raised the ID3 accuracy from 69% to 92%. This result was obtained because of accuracy/rejection plot identified a threshold confidence of 0.66 at which 71% of the patients, classified by highly reliable rules, were accepted. The trade-off was the exclusion from classification of 29% of patients falling in low represented and ambiguous rules. Kaplan-Meier curves showed a significant stratification of stage 4 patients following RO application. In conclusion, we demonstrated that application of the RO improves the classification performance of ID3 decision tree. Stage 4 patients can be stratified in significant groups characterized by the high confidence rules needed for making clinical decisions.