

**Poster 16. PREDICTION OF ONE YEAR MORTALITY IN EXTREMELY PREMATURE NEWBORNS USING CLASSIFICATION TREES**

Ana Januário<sup>1, 2, 3</sup>, Sónia Gouveia<sup>2, 3, 4</sup>, Joaquim Pinto da Costa<sup>3, 5</sup>, Maria Isabel Sá<sup>6</sup>, Alexandra Almeida<sup>7</sup>, Carmen Carvalho<sup>7</sup>, Joaquim Saraiva<sup>6</sup>, Miguel Fonte<sup>7</sup>, Paula Soares<sup>7</sup>

<sup>1</sup>Departamento de Matemática, FC/UP; <sup>2</sup>Gabinete de Estatística, Modelação e Aplicações Computacionais / <sup>3</sup>Centro de Matemática da Universidade do Porto; <sup>4</sup>Instituto de Engenharia Electrónica e Telemática de Aveiro da Universidade de Aveiro; <sup>5</sup>FC/UP; <sup>6</sup>Serviço de Ginecologia e Obstetrícia, MJD/CHP; <sup>7</sup>Serviço de Neonatologia, MJD/CHP.

Centro Hospitalar do Porto (CHP), Porto

Faculdade de Ciências da Universidade do Porto (FC/UP), Porto

Gabinete de Estatística, Modelação e Aplicações Computacionais (GEMAC), Centro de Matemática da Universidade do Porto (CMUP), Porto

Instituto de Engenharia Electrónica e Telemática de Aveiro, Universidade de Aveiro (IEETA/UA), Aveiro

Maternidade Júlio Dinis (MJD), Porto

**Objectives**

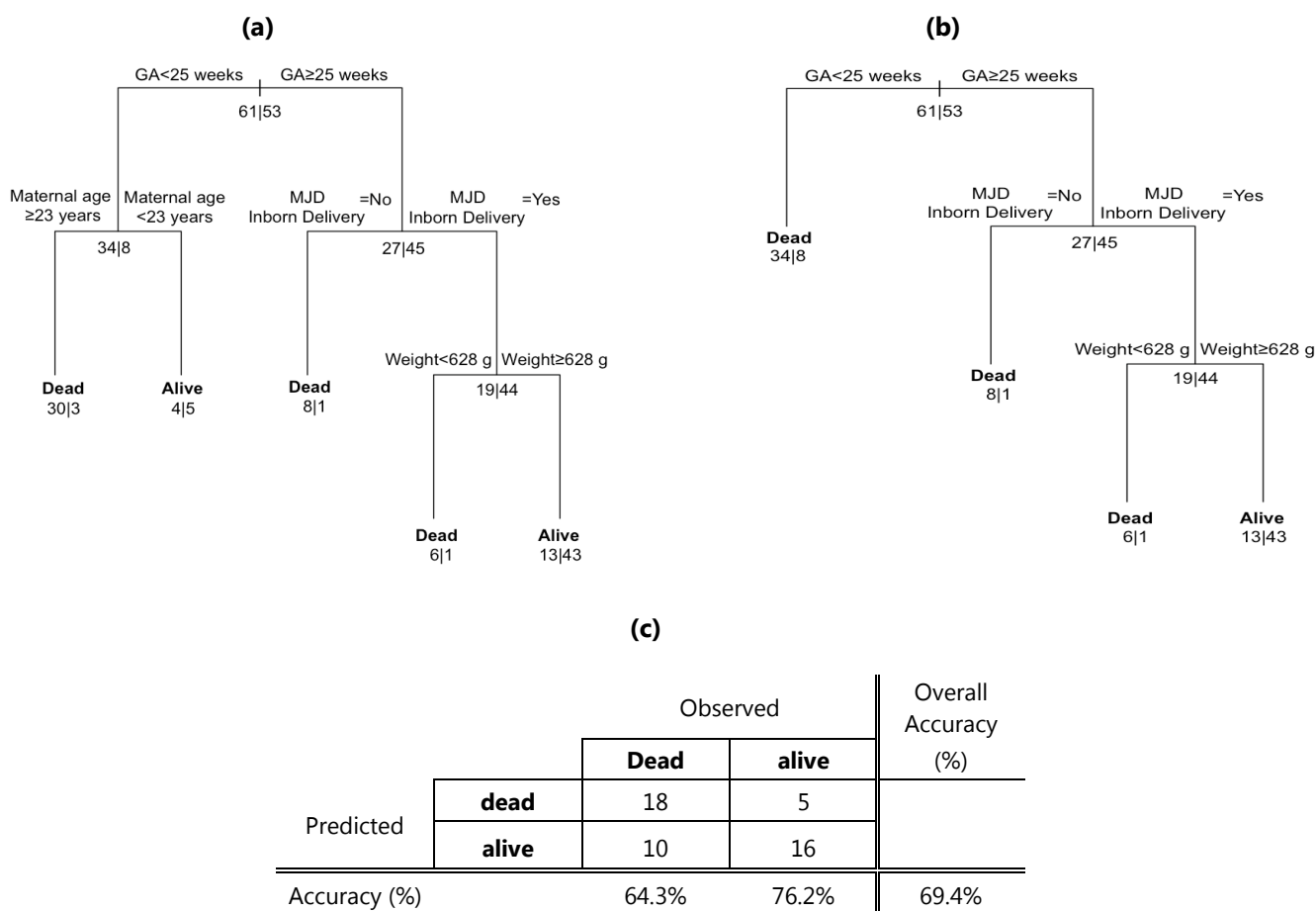
Predictive models are useful tools to help clinical decision and help parental counseling, when dealing with extremely premature newborns. In this work, one year mortality of these infants is predicted with classification trees, which are models easy to interpret that provide a clear and logical representation of the data structure.

**Methods**

Data was collected from 205 newborns (<28 weeks of gestation) followed up at MJD-CHP from 2000 to 2009, comprising one year mortality (yes/no) and 26 variables related to pregnancy and delivery, infant conditions at birth and selected neonatal procedures. There were considered 163 infants, after excluding stillbirths, delivery room death and missing outcomes. The sample was randomly divided into training (70%) and test (30%) sets to construct and validate the predictive model with independent samples, thus avoiding optimistic performance measures. Trees were obtained by recursive partitioning: the root node contains all observations and it is split into 2 nodes and each, by its turn, is divided into 2 other nodes and so on. The variable and cutoff value selected in each node were chosen to obtain a greatest outcome separation. Therefore, the most discriminating variables are typically near the tree top.

**Results**

Figure 1(a) presents the classification tree, highlighting GA, MJD Inborn Delivery, Weight and Maternal age as the most important mortality predictors, with predictive accuracy of 61.2%. The tree was then pruned using cross validation, in order to increase predictive efficiency and reduce classification overfitting to the training set. Figure 1(b) shows that the pruned tree excluded Maternal age and, as reported in Figure 1(c), the overall predictive accuracy increased to 69.4%, with 64.3% and 76.2% of correct predictions in dead/alive outcome. The accuracy of 69.4% for the tree model should be compared with the performance associated with a random prediction, i.e., classify all newborns either as dead or as alive. In the test sample, predictive accuracies for random dead and alive prediction were 57.1% and 42.9%, respectively. The accuracy of the classification tree is higher than that of the random prediction, pointing out the added value of these trees in the prediction of extremely premature newborns.



**Figure 1:** Classification trees to predict mortality outcome of extreme premature newborns before (a) and after pruning (b): e.g., in (a) a newborn with GA  $\leq$  25 weeks, Inborn Delivery at MJD and Weight < 628 g is predicted to died. In each node observed (#dead|#alive) on training sample are provided. Predictive performances were estimated on test sample. Acronyms: GA stands for Gestational age.

### Discussion and Conclusions

The classification tree has important variables for predict mortality: Gestational age, MJD Inborn Delivery and Weight. Our results are agreement with our previous work and literature. The performance of classification tree is higher than that when considering all newborns predicted in the same class.

#### Apresentador

**Ana Januário**, Aluna do Curso de Mestrado em Engenharia Matemática, FC/UP.  
[anafjanuario@hotmail.com](mailto:anafjanuario@hotmail.com)