A MACHINE LEARNING APPROACH FOR PREDICTING PROBABILITY OF DEATH OR DISEASE PROGRESSION IN AN EARLY-TREATED PEDIATRIC AFRICAN COHORT

Sara Domínguez-Rodriguez, Alfredo Tagarro, Miquel Serra Pascual, Kennedy Otworma, Assiya Violari, Sheila Fernández, Tasilla Nhampossa, Maria Lain, Paula Vaz, Ngundu Osebe Behuhuma, Sivapragnarine Danaviah, Elis Dobbelis, Shaun Barnabas, Nicola Cotugno, Paola Zangari, Paolo Palma, Andrea Oletto, Alessandra Nardone, Eleni Nastouli, Moira Spyer, Louise Kuhn, Paolo Rossi, Carlo Giaquinto, Pablo Rojo on behalf of EPIICAL consortium

BACKGROUND
In perinatally HIV infected children, mortality and morbidity are highest in the first months after ART initiation and are linked to advanced disease and late diagnosis. The random forest approach can deal with more predictors than classical models and has no model assumptions such as normality, linearity or hazard proportionality. The aim of this study was to predict the probability of death or clinical progression at a specific time of follow-up.

METHODS

EARTH (EPIICAL consortium) is an African multi-centre cohort enrolling HIV-infected infants treated within 3 months of life
A total of 134 infants with >1 follow-up visit were included in the analysis.
Primary endpoint: Right-censored time to death or clinical progression to AIDS

The model was performed in a training subset (n=95, 70%) and validated on the remaining 30% (n=39)
To predict the outcome, a log-rank random survival forest was performed

RESULTS
A total of 22 infants reached the primary endpoint with 13 (10%) patients dead and 9 (7%) with an AIDS defining condition

The AUC for predicting survival or progression was 0.83, 0.84, and 0.72 for 1-month, 6-months, and 1-year respectively.

Using a machine learning model (random forest), we could accurately classify early treated children according to the probability of death or clinical progression at 1, 6, and 12 months.

This model helps clinicians in the early identification of high risk patients and individualize patient management based on the probability of death or clinical progression to AIDS at time of diagnosis

The most important predictors of reaching the primary endpoint were baseline HIV viral load, age at diagnosis, weight-for-age, baseline CD4 count, time to suppression, and breastfeeding time (Figure 1).

Figure 1. Variable importance of predictors according to the model

In the validation, the model predicted a higher probability of mortality/progression among children who did indeed die or progress to AIDS, as compared to the group of children who did well (1-month: 14% vs. 0.01%, p-value=0.045; 6-months: 62% vs. 0.03%, p-value=0.019; 12-months: 76% vs. 16%, p-value=0.012).

Figure 2. Probability of death/progression