ML Approaches to Paediatric Febrile Illness

Annex C - Data Preprocessing and Outcome Construction Protocol

Outcome Classification Protocol

Spot Sepsis Statistical Analysis Plan Outcome Definitions

$$\boxed{\text{Category I (Severe)}} \qquad \mathrm{I}_i = \begin{cases} 1, \text{ if } (\texttt{Days to Death} \leq 2) \vee (\texttt{Days to Organ Support} \leq 2) \\ \vee\, (\texttt{Discharged Home to Die} \leq 2), \\ \\ 0, \text{ otherwise.} \end{cases}$$

$$\boxed{\text{Category II (Probable Severe)}} \qquad \mathrm{II}_i = \begin{cases} 1, & \text{if } \left[(\texttt{Days to Death}_i > 2 \wedge \texttt{Days to Death}_i \leq 28) \vee (\texttt{Length of Stay}_i > 2) \right] \wedge \mathrm{I}_i = 0, \\ 0, & \text{otherwise.} \end{cases}$$

$$\boxed{\text{Category III (Probable Non-Severe)}} \qquad \mathrm{III}_i = \begin{cases} 1, & \text{if } \left[\left(\texttt{Admitted}_i = 1 \wedge \texttt{Length of Stay}_i \leq 2 \right) \wedge \mathrm{I}_i = 0 \wedge \mathrm{II}_i = 0 \right] \\ & \qquad \vee \left(\texttt{Admitted}_i = 0 \wedge \texttt{Ongoing Symptoms at Day 28}_i = 1 \right), \\ 0, & \text{otherwise.} \end{cases}$$

Four-level Outcome Classification (*new*)

$$\mathrm{Outcome}_i^{(4)} = \begin{cases} \text{Severe (onset } < 24 \text{ h)} & \text{if child classified Severe and } \textit{days to developing severe illness} = 0, \\ \text{Severe (onset } \geq 24 \text{ h)} & \text{if child classified Severe and } \textit{days to developing severe illness} \geq 1, \\ \text{Probable Severe} & \text{if child classified Probable Severe,} \\ \text{Probable Non-Severe} & \text{if child classified Probable Non-Severe.} \end{cases}$$

Machine Learning Grid Search

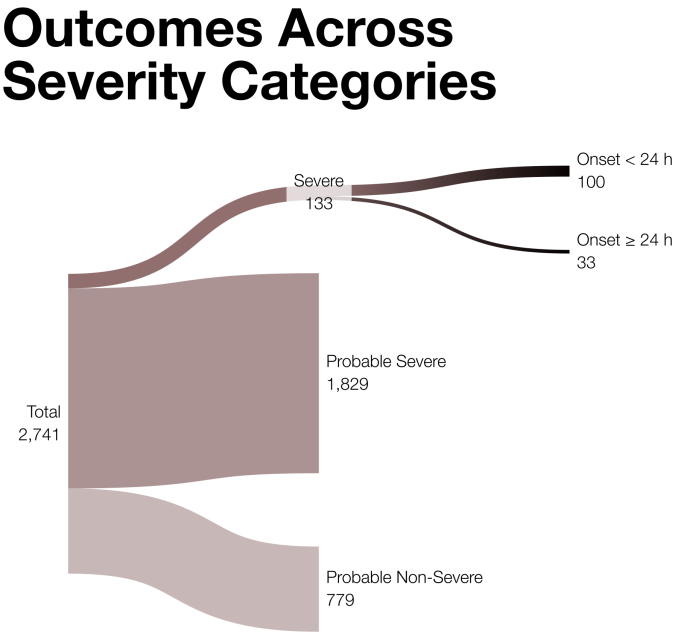

Figure 1: Flow of Outcome Severity Classifications

Model Algorithm	Engine	Tuned Hyperparameter	Grid Type & Size
Decision Tree	`rpart`	`cost_complexity`	Regular (4×4×4), 64 combos
		`tree_depth`	
		`min_n`	
Random Forest	`ranger`	`mtry`	Regular (8×8), 64 combos
		`min_n`	
Gradient Boosting (XGBoost)	`xgboost`	`mtry`	Random, 64 combos
		`min_n`	
		`tree_depth`	
		`learn_rate`	
		`sample_size`	
Support Vector Machine (RBF)	`kernlab`	`cost`	Random, 64 combos
		`rbf_sigma`	
Penalised Regression	`glmnet`	`penalty`	Regular (8×8), 64 combos
		`mixture`	

*The performance of each parameter combination was evaluated using a 5-repeat, 5-fold stratified cross-validation on the training data.