ML Approaches to Paediatric Febrile Illness

Annex D - Meta Learner Architecture

Description of the Meta Learner

The meta-learner approach was examined as a potential avenue to leverage the diverse predictive strengths of the specialist models while mitigating risks of overfitting from highly correlated inputs. To robustly estimate the performance of this strategy and to interrogate its internal mechanics, the entire architecture was constructed and evaluated independently within each of the 25 folds of the 5x5 cross-validation.

Within each fold, the training features for the meta-learner were generated from that fold's unique calibration partition. For each observation, the four specialist models produced a vector of five probability predictions, one for each outcome class, resulting in a 20-dimensional feature set. To circumvent potential issues of multicollinearity, Principal Component Analysis (PCA) was integrated as a preprocessing step. This transformation converted the 20 correlated specialist predictions into a smaller set of orthogonal principal components, with the number of components dynamically determined for each fold by retaining the minimum number required to explain at least 90% of the variance in the original feature set.

A multinomial logistic regression model was then trained on these principal components to serve as the meta-learner for that specific fold. To understand the consistent decision-making logic this strategy employed, the model coefficients and PCA loadings were averaged across all 25 independently trained meta-learners. This analysis revealed a highly targeted and consistent emergent strategy. The model learned to create specialised components to identify specific outcomes rather than operating on broad trade-offs. For instance, one principal component consistently emerged as a dedicated detector for late-onset severe cases, being primarily composed of severe signals from the base models. Another component served to unambiguously isolate 'Probable Non-Severe' cases. This systematic approach demonstrated that the meta-learner had learned to construct and use a series of specific, targeted signals to classify cases.

The consistent and effective strategy observed across the cross-validation runs confirmed the viability of the meta-learner approach. This justified the subsequent development of a single, definitive meta-learner, trained on the entire available training dataset, for the final deployed cascade model.

Interpreting the Meta Learner

Table 1: Summary of Averaged Meta-Learner Strategy Across All Runs

		Average Coefficient (Log-Odds) for Outcome Level			
Principal Component	Component Definition (Top Terms)	Non-Severe	Onset greater than 24 hours	Probable Non-Severe	Probable Severe
(Intercept)	NA	4.47	-22.30	5.52	6.41
PC1	Non-Severe (32%) | Probable Severe (28%) | Onset within 24 hours (20%) | Onset greater than 24 hours (17%) | Probable Non-Severe (3%)	2.03	-6.17	1.51	1.25
PC2	Onset greater than 24 hours (32%) | Onset within 24 hours (29%) | Probable Severe (29%) | Non-Severe (8%) | Probable Non-Severe (2%)	0.13	2.83	0.12	0.10
PC3	Probable Non-Severe (78%) | Non-Severe (21%) | Probable Severe (1%) | Onset within 24 hours (0%) | Onset greater than 24 hours (0%)	-0.24	4.27	-0.19	-0.22
PC4	Onset greater than 24 hours (51%) | Onset within 24 hours (49%) | Probable Severe (0%) | Non-Severe (0%) | Probable Non-Severe (0%)	-0.03	0.69	-0.09	-0.05

*Coefficients are presented as average log-odds effects of each principal component (PC) on the likelihood of classification into each outcome category, relative to the baseline "Onset within 24 hours".
Positive values indicate an increased likelihood of classification into the given outcome when the PC score is high; negative values indicate a decreased likelihood.
*Estimates are averaged across 25 stratified cross-validation folds from a multinomial logistic regression meta-model.

The intercept reveals the model's baseline prediction for an "average" case before any principal component evidence is considered. In this analysis, **`Onset within 24 hours`** serves as the reference category, so its intercept is, by definition, **zero**. The intercepts for all other categories represent their intrinsic log-odds of occurring relative to this zero-point.

The model's strongest initial assumption is a **`Probable Severe`** outcome, which has the highest intercept at **+6.41**, making it significantly more likely than the baseline. This is followed closely by **`Probable Non-Severe`**(**+5.52**) and **`Non-Severe`** (**+4.47**). In stark contrast, the **`Onset greater than 24 hours`** category has a massive negative intercept of **-22.30**, indicating it is intrinsically far less probable than the `Onset within 24 hours` baseline.

Taken together, the intercepts establish the model's starting point: it assumes an "average" case is most likely to be `Probable Severe`, and that a late-onset severe event is extremely rare. The principal components then provide the specific evidence required to shift a prediction away from these initial odds.

Principal Component 1 acts as a broad signal separator rather than a simple detector for a single outcome. It is composed of a diverse mix of specialist predictions, primarily 'Non-Severe' (32%), 'Probable Severe' (28%), and 'Onset within 24 hours' (20%). The meta-learner leverages this mixed signal to create a clear distinction: a one-unit rise in the PC1 score yields a +2.03 log-odds shift toward a 'Non-Severe' classification but, more powerfully, a **-6.17** log-odds shift away from 'Onset > 24 hours'. Although 'Probable Severe' predictions account for over a quarter of PC1's variance, the resulting effect on a 'Probable Severe' classification is modest (+1.25 log-odds). This contrast illustrates how the meta-learner uses a component representing a consensus of common outcomes to primarily rule out a specific, critical diagnosis.

Principal Component 2 encapsulates a broad "General Severity" signal, loading almost equally on specialist predictions for 'Onset > 24 hours' (**32%**), 'Onset within 24 hours' (**29%**), and 'Probable Severe' (**29%**). The meta-learner utilizes this component as a powerful and specific trigger for a late-onset diagnosis; a one-unit increase in the PC2 score corresponds to a **+2.83** log-odds shift toward the 'Onset > 24 hours' outcome, relative to the baseline. Notably, this component has a negligible effect on the log-odds of any other outcome category. This demonstrates the model's ability to synthesize a heterogeneous set of high-risk signals into a single, focused predictor for one specific severe class.

PC3 is defined almost entirely by non-severe signals ('Probable Non-Severe' at 78% and 'Non-Severe' at 21%). Counter-intuitively, the model uses this component as its single most powerful predictor *for* a late-onset severe diagnosis. A high score on PC3 corresponds to a massive **+4.27** log-odds shift toward the ‘Onset > 24 hours’ outcome. This suggests the meta-learner has discovered a paradoxical rule: a specific pattern of specialist models strongly agreeing on a *non-severe* outcome is, in fact, the most powerful available signal that the case will ultimately become a late-onset severe event.

PC4 is a highly specialised component driven exclusively by the interplay between the two main severe outcomes: ‘Onset > 24 hours’ (51%) and ‘Onset within 24 hours’ (49%). It functions as a final, fine-tuning mechanism. A high score on PC4 provides a modest **+0.69** log-odds push toward an ‘Onset > 24 hours’ classification over the baseline. After the other components have made their primary determinations, PC4 helps to resolve ambiguity between the two most critical classes.