	Additional layers	Units	Learning rate	Dropout rate	Batch size	Epochs	L2 penalty	Activation function	Optimizer
Stage 1	None	None	0.0016	0.364	110	46	0.0039	SeLu	SGD
Stage 2	None	None	0.0054	0.349	21	50	0.00612	SeLu	RMSProp
Stage 3	2	86	0.01	0 *	71	86	0.000001	SeLu	RMSProp
Flat	None	None	0.01	0.5	84	88	0.01	SeLu	SGD

Table *Errore. Nel documento non esiste testo dello stile specificato.**.1: Results of the Bayesian optimization.** The table reports, for each stage of the hierarchical model and for the flat multiclass classifier, the architecture selected during Bayesian optimization (including the number of hidden layers and neurons), activation functions, regularization parameters (dropout rate and L2 penalty), optimization algorithms, learning rates, batch sizes, and number of training epochs. These values reflect the best-performing configurations identified within the predefined search space. Epoch counts and learning rates correspond to the initial settings provided to the training loop but may have been dynamically modified by early stopping and learning rate scheduling callbacks during model training. * The dropout rate was manually changed to 0.2 to avoid overfitting.*

	Baseline	CV	Testing	Balanced batching	PIDgeon
Stage 1	0.869 ± 0.025	0.890 ± 0.016	0.912 ± 0.012	0.880 ± 0.007	0.961
Stage 2	0.830 ± 0.001	0.838 ± 0.002	0.599 ± 0.024	0.570 ± 0.024	0.875
Stage 3	0.427 ± 0.016	0.523 ± 0.012	0.335 ± 0.029	0.328 ± 0.014	0.374
Flat	0.505 ± 0.009	0.540 ± 0.012	0.374 ± 0.008	0.369 ± 0.008	0.480

***Table** Errore. Nel documento non esiste testo dello stile specificato**..2: Comparison of performances.** This table shows the average balanced accuracies calculated from the aggregated confusion matrices for each setup with standard error. "Baseline" refers to the models before Bayesian optimization of the hyperparameters, "CV" refers to cross-validation, and "Balanced batching" refers to a second round of testing carried on using a balanced batch generator instead of class weights. For the performances of the PIDgeon classifier, no standard deviation or standard error was provided.*