MFT-MCH ML Track Matching

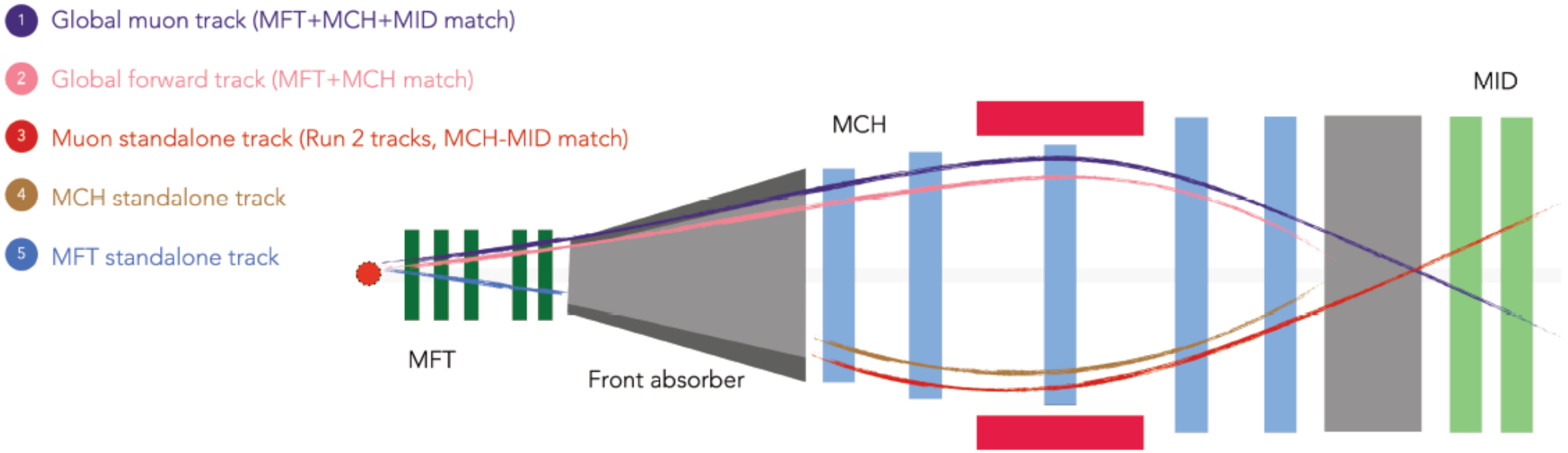


FIGURE 4.10: The different track types present in the muon spectrometer for Run 3.

Model Input params

Input params:

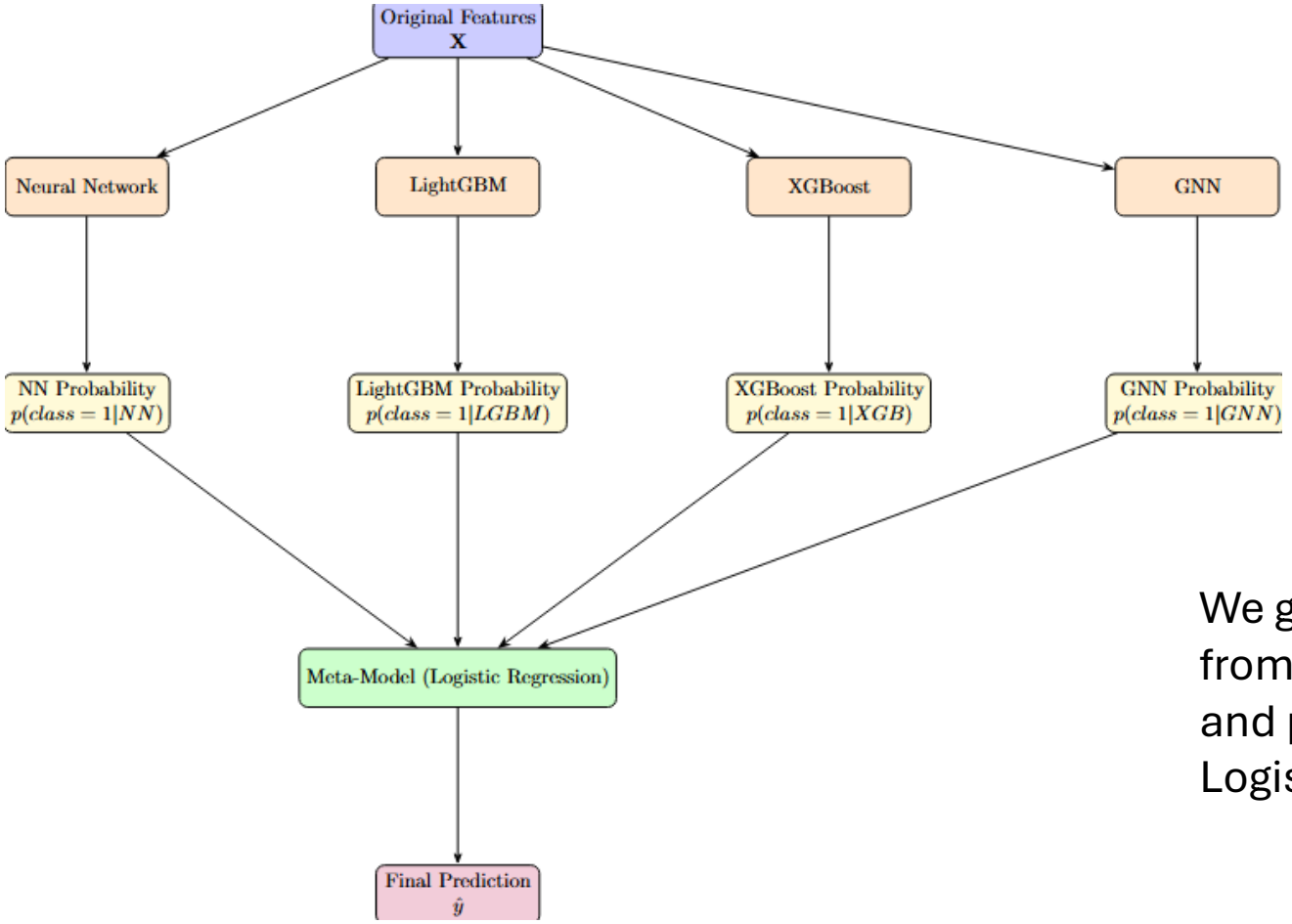
'X_MCH', 'Y_MCH', 'phi_MCH', 'tanL_MCH', 'invqpt_MCH',
'X_MFT', 'Y_MFT', 'phi_MFT', 'tanL_MFT', 'invqpt_MFT',
'chi2'

Model

Output:

For each MCH Track, we have 5 choices for a MFT track to match with. Model outputs a classification choice for the MFT track. (Before this, model would have simply said yes or no to each MFT-MCH pair, where the pair was based on the highest chi2 MFT track)

Combined (Stacked) Model



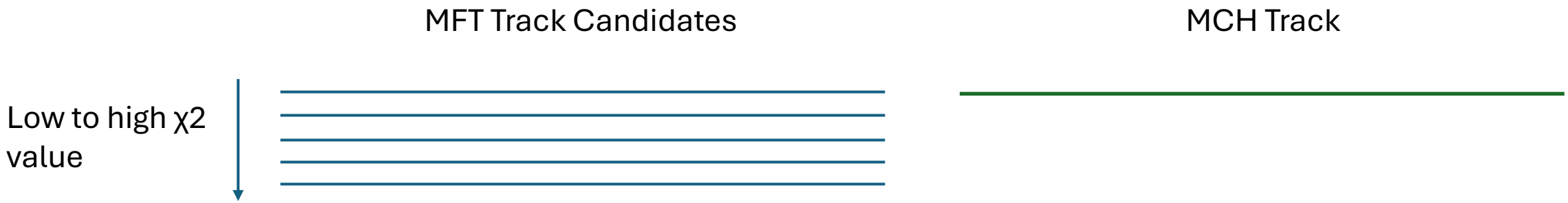
We get a new learned representation from combining predictions of all models and putting these predictions through a Logistic regression

Model Results

Model	Train F1	Test F1	Train AUC	Test AUC	SBR	FPR	PR AUC
NN	0.958	0.956	0.997	0.997	0.931	0.011	0.973
GNN	0.937	0.936	0.997	0.996	0.907	0.016	0.968
XGBoost	0.963	0.956	0.998	0.997	0.942	0.009	0.975
LightGBM	0.962	0.955	0.998	0.997	0.940	0.010	0.974
Stacked	0.965	0.960	0.998	0.998	0.953	0.007	0.976

- The Stacked model seems to be outperforming all the other models (as expected)

Only match MFT candidate with lowest χ2 value and comparing with stacked model:



- χ2 model only uses the lowest MFT track. Yes or no if we have a correct match.
- ML model uses (here at least) 5 candidates and chooses between the 5 candidates which is a match. Choose tracks based on many track parameters fed into ML (including χ2)

Only match MFT candidate with lowest χ2 value and comparing with stacked model:

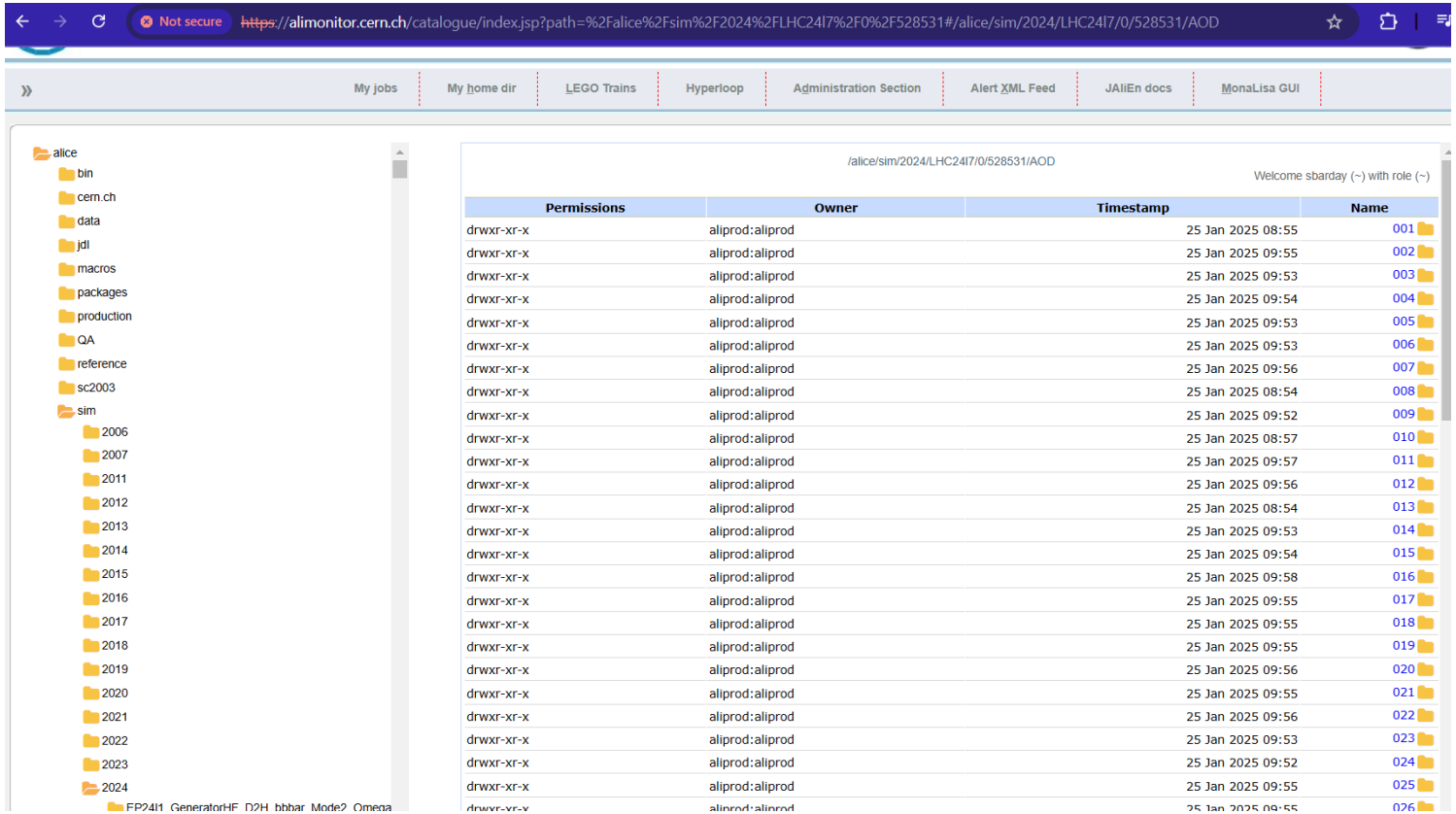
Method	F1 Score	SBR	FPR
Lowest χ^2 Candidate	0.795	0.664	0.078
Stacked Model	0.960	0.953	0.007

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

(Precision/Purity or Signal to background ratio)

- The ML classifier boosts F1 by roughly 0.165, which is over a 20% relative improvement
- The lowest-χ2 rule yields a lot of false positives (only about 66% of its “match” predictions are correct), whereas the stacked model pushes SBR above 95%.
- χ2: FPR ≈ 0.078 (i.e. about 7.8% of non-matches are mislabeled).
  Stacked Model: FPR ≈ 0.007 (fewer than 1% of non-matches slip through).
  ML reduces the false-positive rate by more than an order of magnitude.

Data Used



https://its.cern.ch/jira/browse/O2-5641

Data Description

Description

1 J/ψ + 1 ψ(2S) generated at forward rapidity and injected in a MB pp event every 5 events

Anchor period: 2022 pass7

Special requests:

- Remap of the MID efficiencies (see O2-5613)
- modification of the MFT matching workflow applying these conditions: **o2-globalfwd-matcher-workflow --configKeyValues "FwdMatching.saveMode=3;FwdMatching.nCandidates=5**

Link to the grid folder with config and JDL:
https://alimonitor.cern.ch/catalogue/#/alice/cern.ch/user/l/lmichele/selfjobs/TestPromptCharmoniaAnchor22pass7_withoutMidEff

ini file: O2DPG/MC/config/PWGDQ/ini/Generator_InjectedPromptCharmoniaFwdy_TriggerGap.ini

- Estimation based on previous requests O2-4884
- Events: 100M
- Disk space: ~11.5 TB
- Duration: ~4.5 days at 10K CPUs

Performance Metrics

	$\mathrm{ML} = 0$	$\mathrm{ML} = 1$
$\mathrm{MC} = 0$	TN	FP
$\mathrm{MC} = 1$	FN	TP

TABLE 6.1: Table describing the variables TN, FN, TP and TN.

$$\text{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

(Precision or Signal to background ratio)

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$