Deep Learning Studies for the Measurement of the Top Quark Mass

David Okpaga

MSc Defense

September 29, 2025

Supervisors

Dr. Andrea Helen Knue

Dr. Javier Jimenez Pena

technische universität dortmund

UNIVERSITÉ Clermont Auvergne

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

IMAPP

EU

ATLAS EXPERIMENT

“Are you sitting comfortably?
Then we’ll begin...”

Measurement Idea - Baseline approach

- **Why the top quark mass matters**
- Heaviest known fundamental particle – $\mathbf{m_{top}}$ plays a key role in:
  - Electroweak precision tests
  - Stability of the SM vacuum
  - Connection to Higgs boson physics
- The goal is to measure m_{top} by reconstructing top decays in the **lepton + jets** channel.
- Using two kinematic observables:
  - $M_w^{reco} = m(q_1, q_2)$
  - $M_{lb}^{reco} = m(b_{lep}, lep)$
- **Baseline method: Kinematic Likelihood Fit (KLFitter)**
- Assigns jets to partons using likelihood maximization.
- Achieves ~**63% correct matching efficiency**.
- Provides a physics-motivated benchmark, but still limited by combinatorial background.

Why go beyond KLFitter?

- **Limitations of KLFitter**
  - Efficiency capped at ~63%
  - Sensitive to resolution effects and wrong jet assignments
  - Struggles with light-quark assignment from W decays (low-p_T jets often mis-assigned)

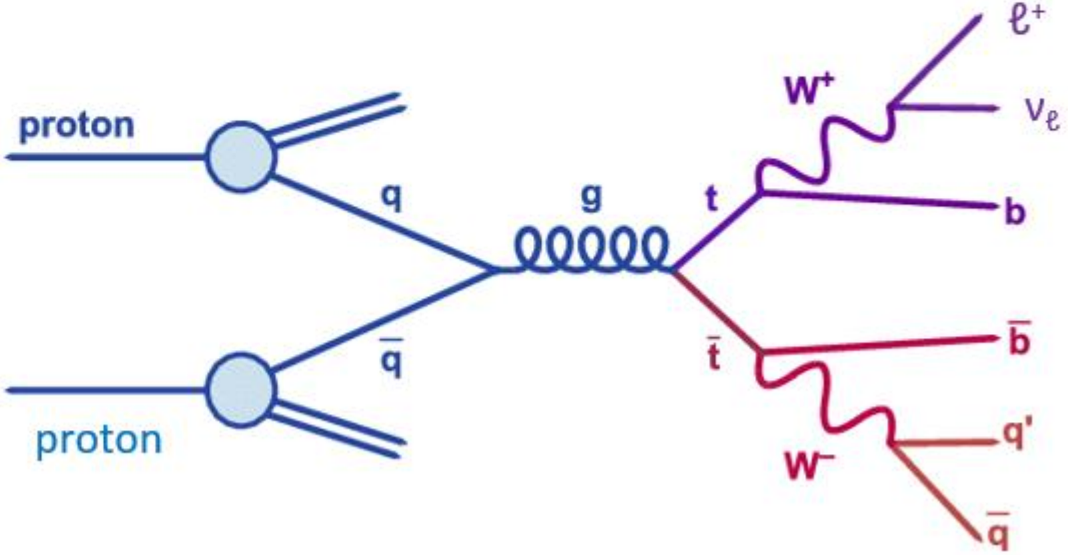
- **Impact of limitations**
  - Lower purity in reconstructed observables
  - Reduced statistical precision in mass measurement
  - Bias from incorrect jet-parton permutations

These limitations motivate more flexible, data-driven methods to capture complex correlations — leading to machine learning approaches (DNN).

Production & Decay

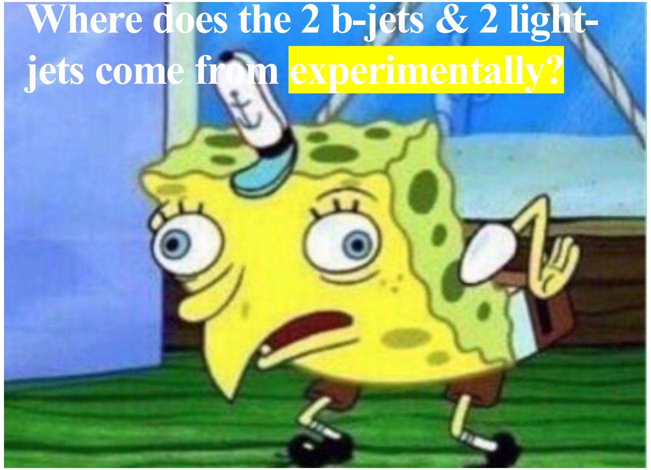
Production

Decay



- Top quark pair are produced in the largest cross-section of all top processes.
- We expect about 51 million ttbar events in the lepton+jets channel in the dataset (before selection)

- 4 Quarks → 4 Jets
- 4 quarks hadronize into **jets**:
  - 2 b-jets (tagged)
  - 2 light jets (from W decays)
- 1 lepton
- 1 neutrino inferred via missing E_T



Jet-Parton reconstruction

The Two Observables:

Option 1: $(M_t^{reco})^2 = | p_b + p_q + p_{q'} |^2$ → Direct sensitivity but high uncertainty

- Reconstruct the top mass from the **b-jet + two light jets**
- Uses 3 jets

❖ **Advantage** → Directly reconstruct hadronic top mass

Problem!

- Jets have large experimental uncertainties:
  - Jet energy scale uncertainty
  - Jet energy resolution
- M_{top} measurement has larger uncertainty

Option 2: $(M_{b\ell}^{reco})^2 = | p_b + p_\ell |^2$ → Indirect sensitivity but lower uncertainty

- Reconstruct the top mass from the **b-jet + lepton**
- Uses 1 jet

❖ **Advantage** → lepton momentum has much smaller uncertainty than jets
→ m_{top} measurement has smaller uncertainty

Problem!

- Also depends on correctly assigning the **b-jet and lepton pair** (combinatorial problem).

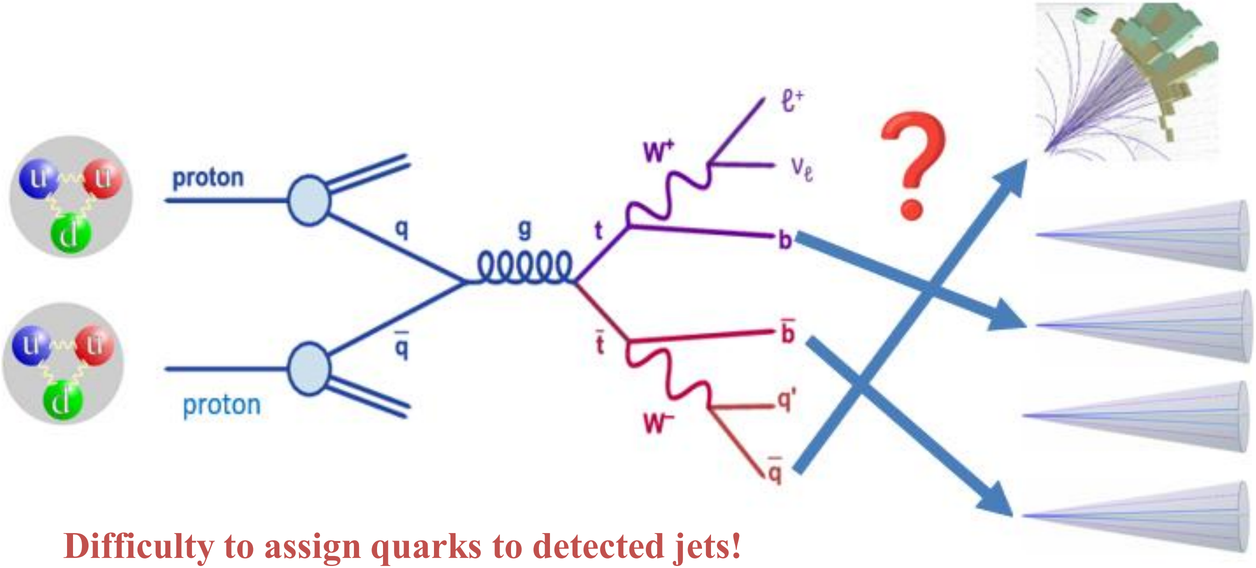
Jet-To-Quark Assignment Classes (Jet mapping)

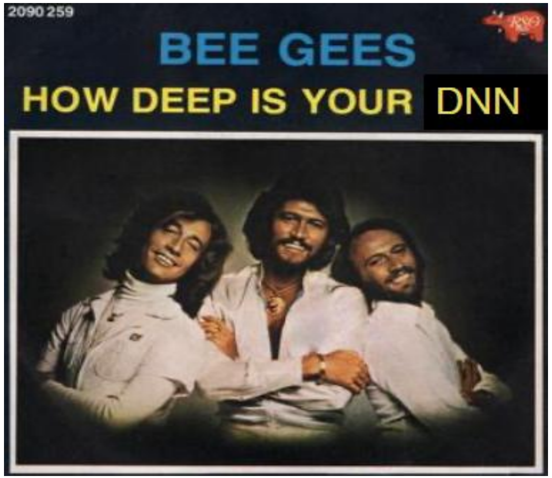
S/N	b_{had}	q_1	q_2	b_{lep}
1	Jet1	Jet2	Jet3	Jet4
2	Jet1	Jet2	Jet4	Jet3
3	Jet1	Jet3	Jet4	Jet2
4	Jet2	Jet1	Jet3	Jet4

Output Y:

- 24 possible permutations
- 12 valid permutations of 4 jets assigned to 4 quarks (after W jet symmetry)
- In reality, we have less than 12 permutations if only b-jets are allowed in the position of b-quarks.

Combinatorial background from wrong jet-parton assignments



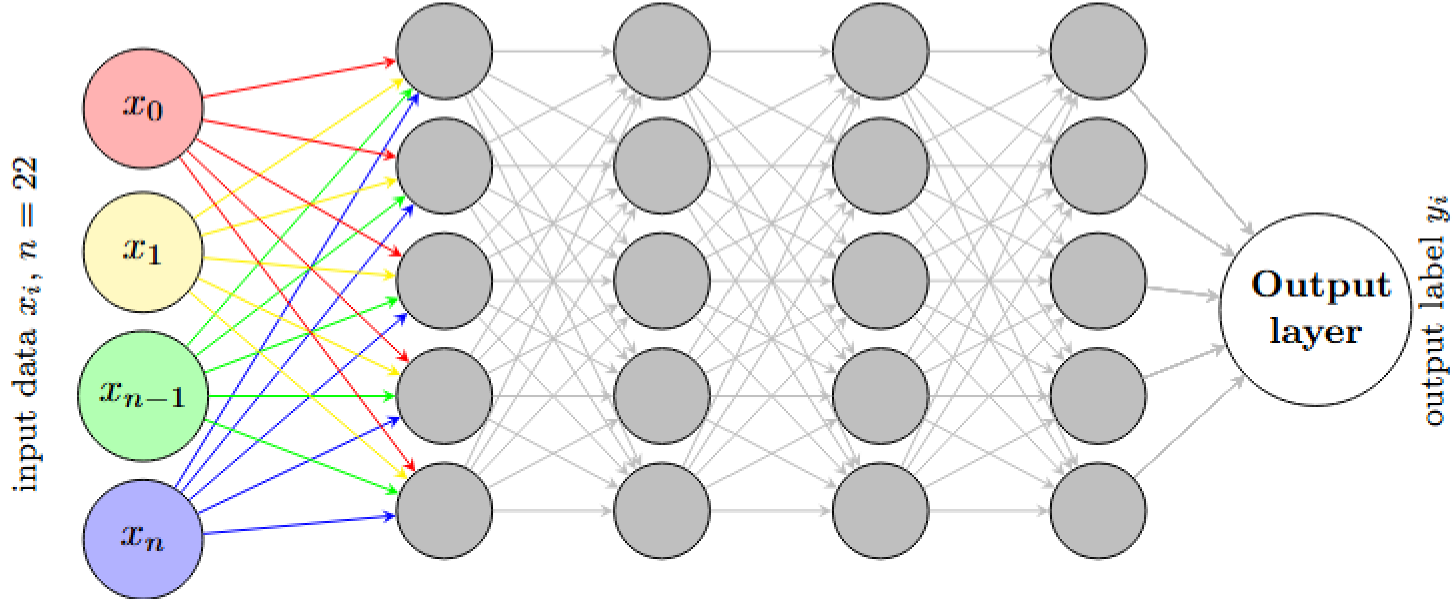


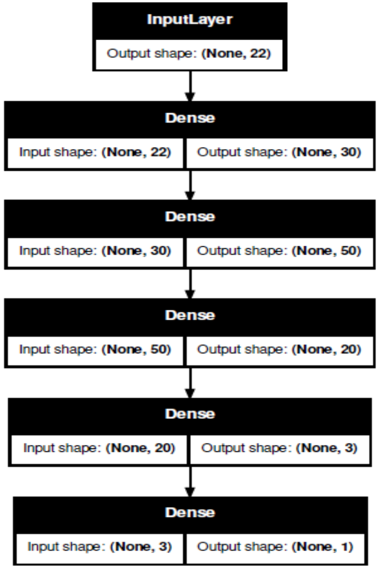
Difficulty to assign quarks to detected jets!

Now, the **main problem**:

- To compute the **observables**, we must assign which jet corresponds to which **quark**:
  - Which jet is **b**?
  - Which jets are **q and $\bar{q}'$**?

DNN Architecture





- Signal: correct permutation, background: all wrong permutations
- Input variables: four momenta of lepton and jets and missing E_T
- For each event, all permutations are evaluated and the one obtaining the highest DNN score, DNN_{High}, is selected.

Event selection

Event selection criteria:

- Single-lepton channel: exactly one isolated electron or muon.
- Events required to pass a single-lepton trigger (electron or muon + jets).
- At least one reconstructed primary vertex.
- At least four reconstructed jets with $p_T > 25\ GeV$ and $|\ \eta\ | < 2.5$.
- At least two jets must be b-tagged.

Jet-Parton matching

Matching criteria:

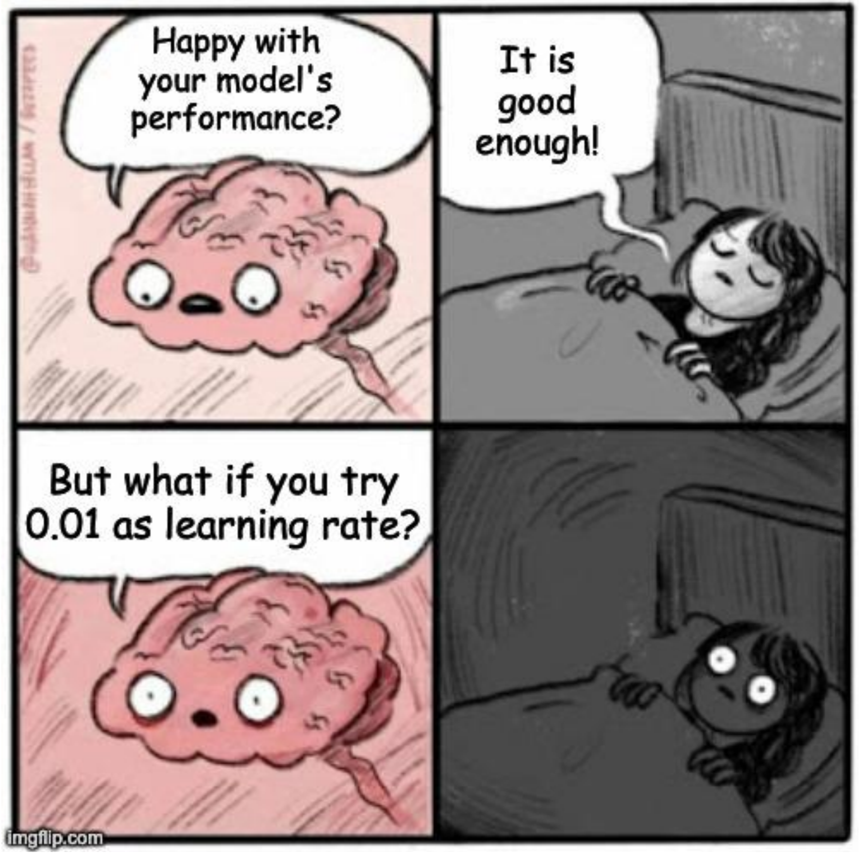
- Only **b-tagged jets** can be assigned to **b-quark positions** (i.e., b_{had}, b_{lep})
- Only **untagged jets** are allowed for **light-quark positions** (i.e., q_1, q_2) from hadronic W decay
- Matching criterion: if ΔR (parton, jet) < 0.3: consider jet as matched
- Matching criterion for event: all four partons are matched to a jet

Why use $p_T > 25\ GeV$ for selection:

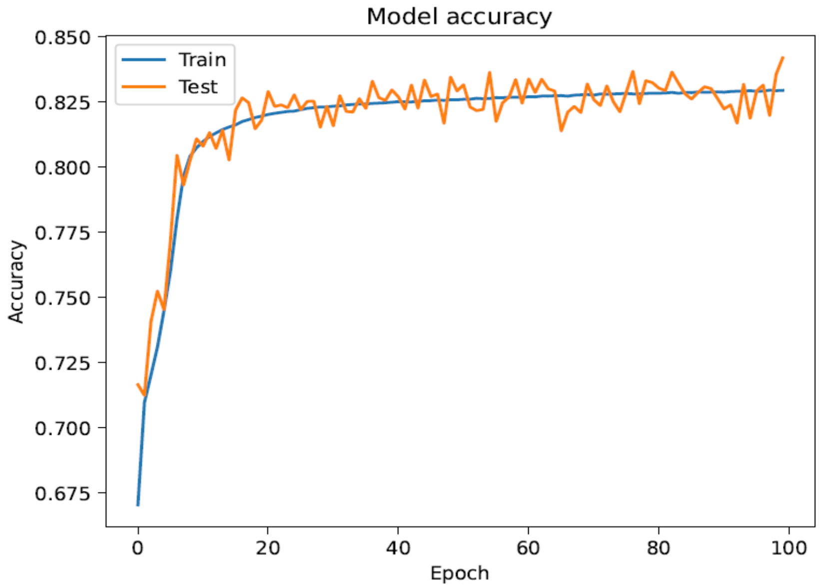
- **Jets with higher p_T have smaller uncertainties** and are more likely to be well-reconstructed.
- For lower p_T, jets are harder to match accurately due to increased detector **uncertainties**.
- For a jet with p_T (J3) $= 24.5\ GeV$, it **does not** meet the $p_T > 25\ GeV$ requirement
  - **J3** missing → lq2 does not have a matched jet (lq2 is unmatched).
- The training takes 2 days and it corresponds to the 2017 dataset and I used simulation for 172.5 GeV
- The training was done for 5 mass points (171 GeV, 172 GeV, 172.5 GeV, 173 GeV, 174 GeV)

Optimization features

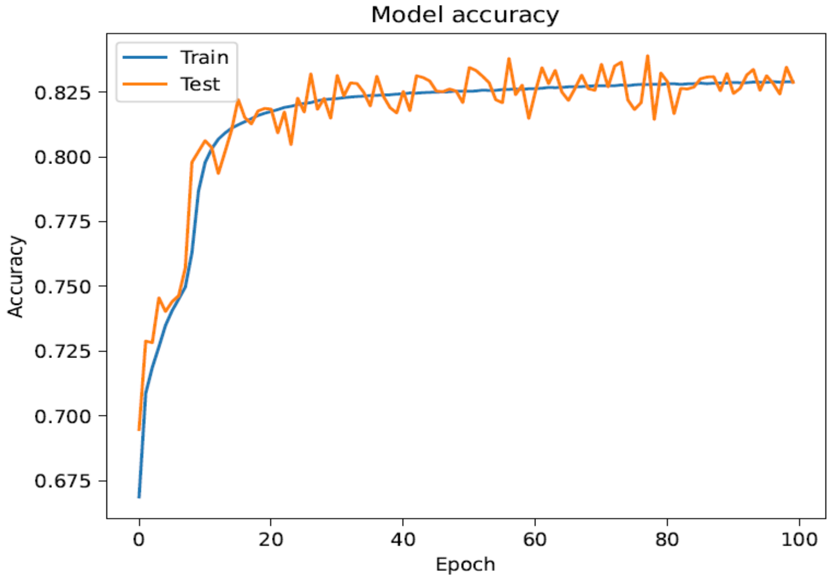
- Number of nodes in input layer: 22 variables
- Number of nodes in hidden layers: 30, 50, 20, 3
- Output layer: 1
- N_{batch}: 5000
- N_{epoch}: 100
- Optimisation algorithm: ADAM
- Learning rate: 0.005
- Loss function: Binary cross-entropy
- Activation function (hidden layer): ReLU
- Activation function (output layers): Sigmoid



First DNN training: Accuracy



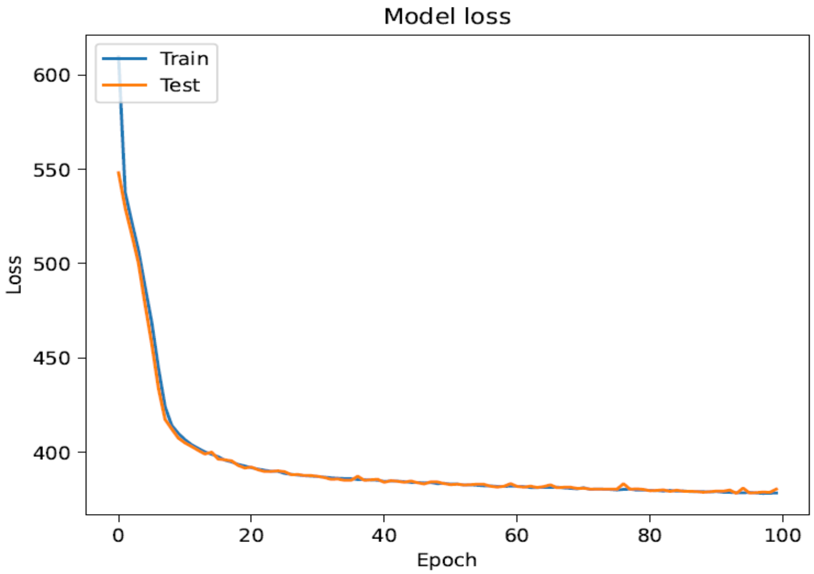
Even



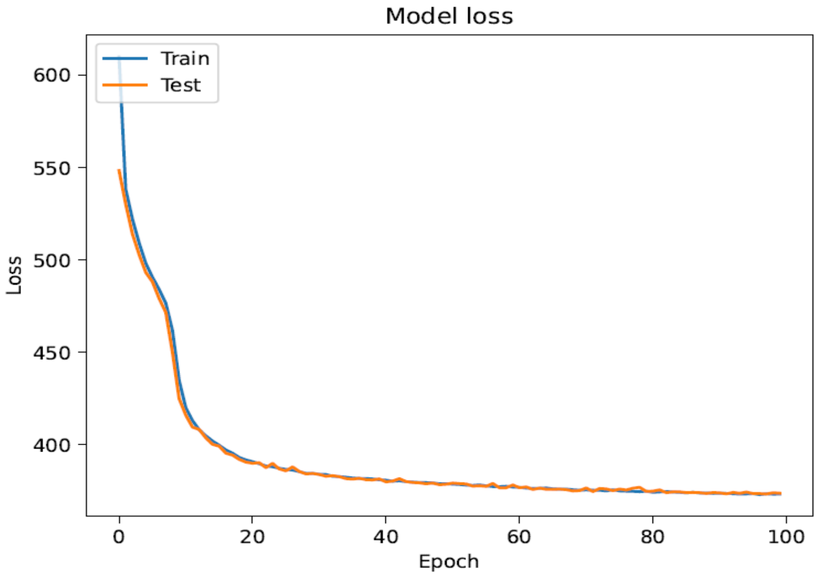
Odd

- DNN reaches 82–83% accuracy after ~30 epochs
- Both training and test accuracy curves are closely aligned
- Very small gap between train and test accuracy → excellent generalization
- No signs of overfitting

First DNN training: Model loss



Even



Odd

- Both show a smooth exponential decay – from ~600+ down to below ~370
- Train and Test Loss are almost identical – overlap closely
- ○ Very good generalization
- ○ No overfitting
- ○ Stable convergence

DNN matching performance at m_{top} = 172.5 GeV

The matching efficiency is defined as the fraction of matchable events among all selected events:

$$\epsilon_{matching} = \frac{N_{matchable}}{N_{matchable} + N_{unmatchable}} = \frac{N_c + N_i}{N_c + N_i + N_u}$$

The reconstruction efficiency is the fraction of correctly matched events among all matched events:

$$\epsilon_{cm} = \frac{N_c}{N_c + N_i}$$

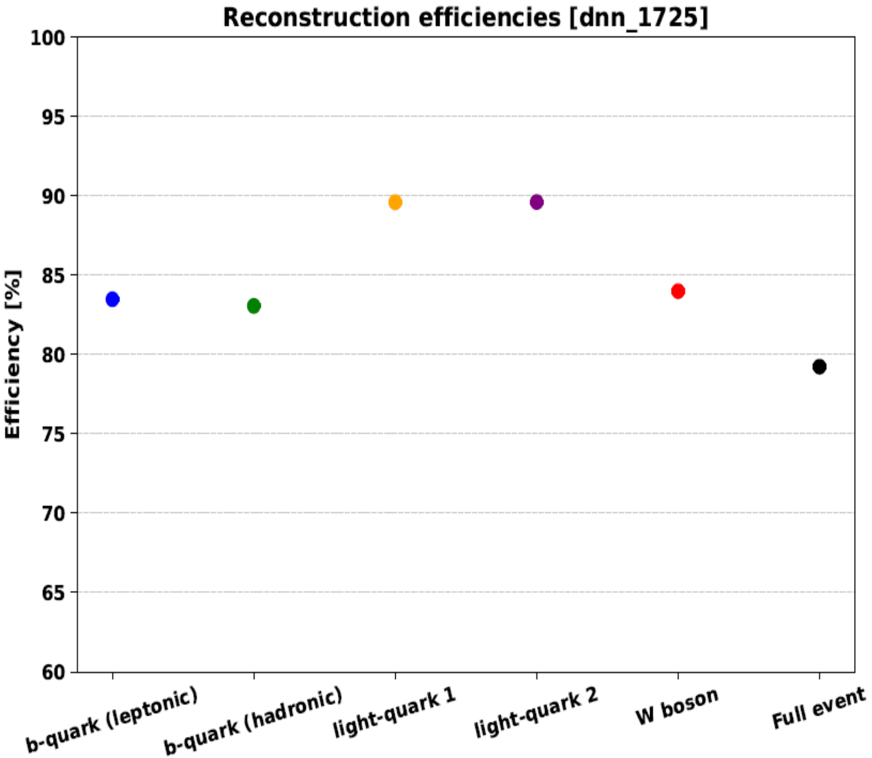
The selection purity is the fraction of correctly matched events among all selected events, regardless of matchability:

$$\pi_{cm} = \frac{N_c}{N_c + N_i + N_u}$$

- Low matching efficiency in $t\bar{t}$ events

  → Caused by low-p_T jets from W boson decays failing selection.

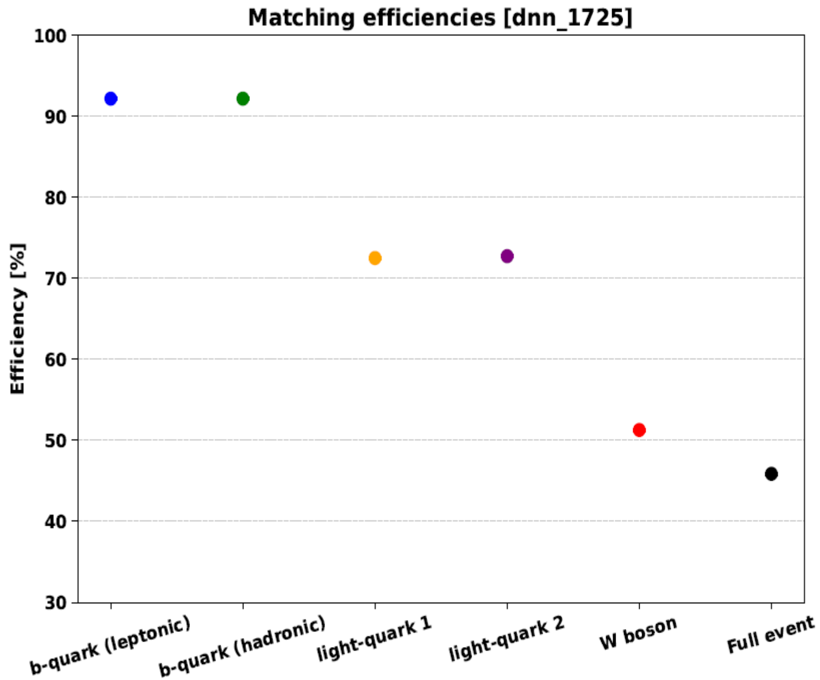
First reconstruction efficiency

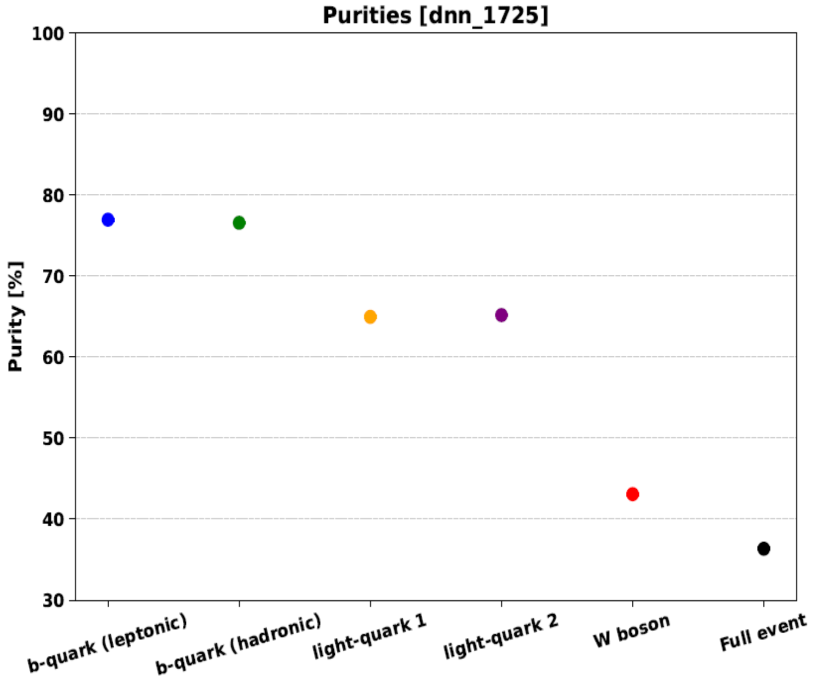


Reconstruction efficiency: Higher for light quarks (~89%) compared to b-quarks (~83%) and full event ~79%.

- Due to simpler kinematics, light quarks are easier to reconstruct.

Matching efficiency and purity





Matching efficiency: High for b-quarks (~92%), moderate for light quarks (~72%).

- **b-jets** are more accurately matched due to **b-tagging**.

Purity: High for b-quarks (~77%) due to **better matching** & **less background**, lower for light quarks (~65%).

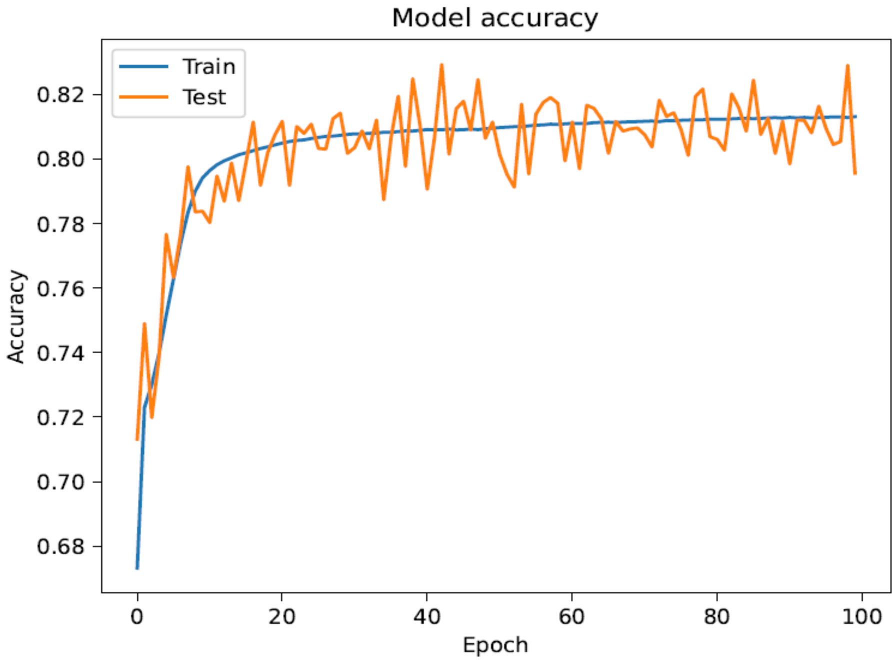
The impact of P_T > 20 GeV on DNN performance

If I lower P_T:

- Do I get more pileup? High prob. of more pileup contamination
- Do I get larger JES uncertainty? Highly possible
- Higher event yield → less clean jets
- Poorer energy resolution
- Worse b-tagging performance
- Longer training, more permutations

○ More jets = More combinations = More permutations

(e.g., using 6 jets instead of 4).



Effect of p_T cuts on efficiency and purity

Quantity	$p_T > 25$ GeV	$p_T > 20$ GeV
Overall metrics		
Matching efficiency (%)	45.88 ± 0.05	48.29 ± 0.39
Reconstruction efficiency (%)	79.24 ± 0.06	71.90 ± 0.50
Purity (%)	36.35 ± 0.05	34.72 ± 0.37
Reconstruction efficiencies (%)		
b_{lep}	80.30 ± 0.04	81.35 ± 0.31
b_{had}	79.43 ± 0.05	81.52 ± 0.31
lq_1	89.08 ± 0.04	84.85 ± 0.32
lq_2	89.07 ± 0.04	85.22 ± 0.32
m_W	83.09 ± 0.06	76.88 ± 0.44
Matching efficiencies (%)		
b_{lep}	92.19 ± 0.03	91.39 ± 0.22
b_{had}	92.21 ± 0.03	91.90 ± 0.21
lq_1	72.56 ± 0.05	74.60 ± 0.34
lq_2	72.79 ± 0.05	74.53 ± 0.34
m_W	51.36 ± 0.05	54.91 ± 0.38
Purities (%)		
b_{lep}	74.02 ± 0.05	74.35 ± 0.34
b_{had}	73.24 ± 0.05	74.91 ± 0.33
lq_1	64.64 ± 0.05	63.30 ± 0.37
lq_2	64.83 ± 0.05	63.51 ± 0.37
m_W	42.67 ± 0.05	42.21 ± 0.38

❖ **Reconstruction Efficiencies:**

- Increase for b_{lep}, b_{had} at $p_T > 20\ GeV$ due to more signal events at lower threshold
- Decrease for light quarks due to more low p_T jets that are harder to match correctly.

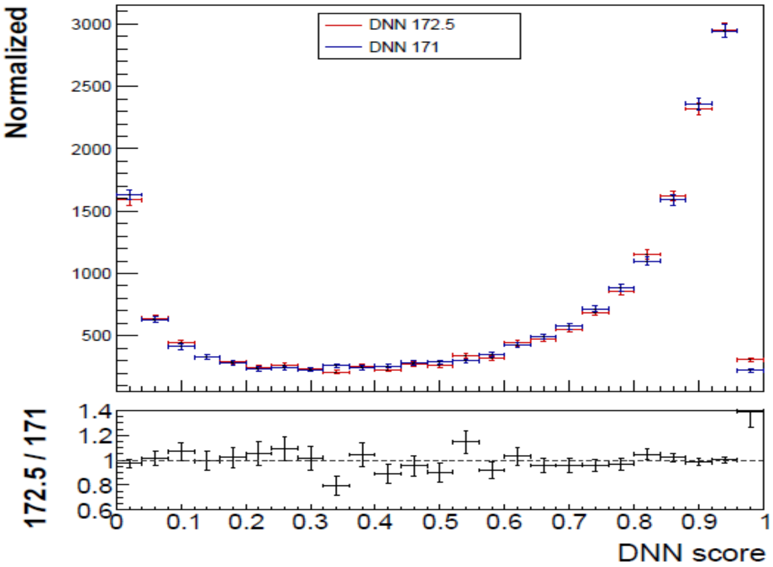
❖ **Matching Efficiencies:**

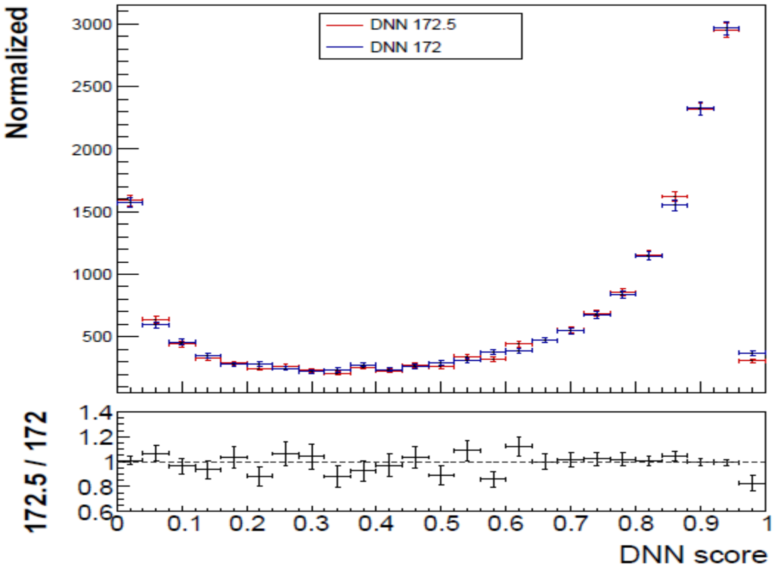
- Increase for $lq_1,\ lq_2$ at $p_T > 20\ GeV$ due to more inclusive jet reconstruction.
- Decrease for m_W and b_{lep} due to background contamination at lower p_T.

❖ **Purity:**

- Increase for b_{lep}, b_{had} at $p_T > 20\ GeV$ because lower threshold allows more high-quality b-jet events to be selected, improving signal purity.
- Decrease for other variables (light quarks, m_W (due to increased background contamination as more low p_T events are included.

Comparison of m_t hypotheses in the DNN output



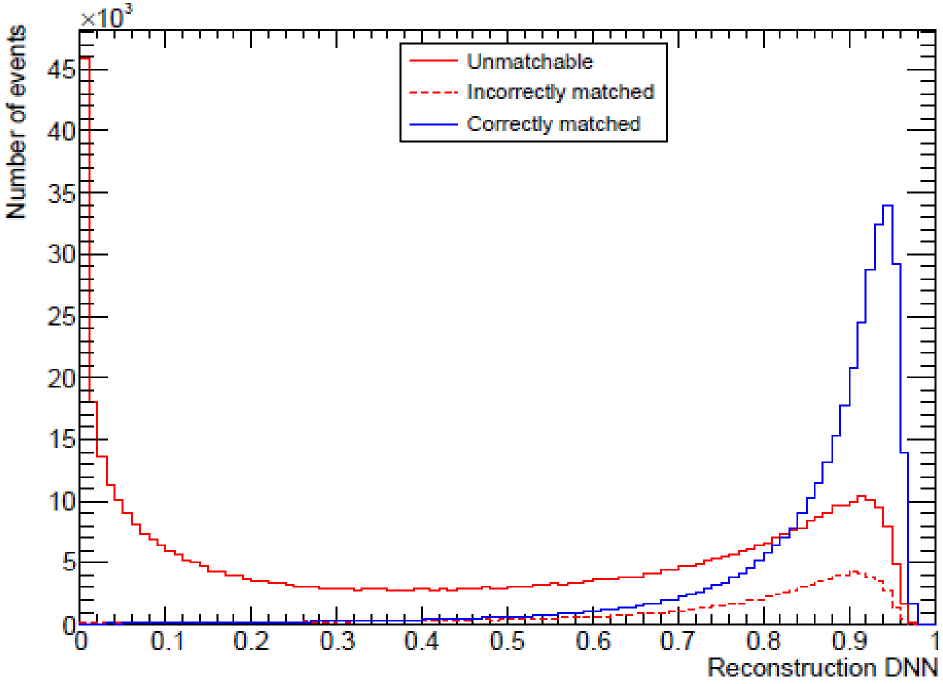


DNN score comparison: Both mass hypotheses show clear peaks as the DNN score increases.

Ratio plots:

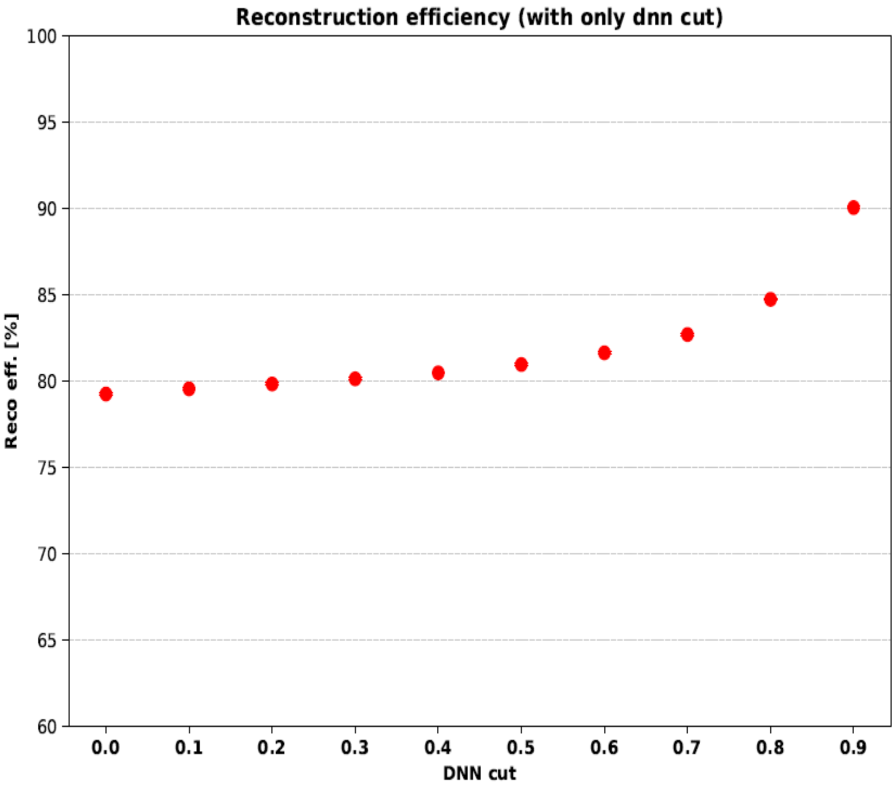
- Ratios (172.5/171 and 172.5/172) stay near 1, indicating minimal difference at high DNN scores.

Reconstruction DNN Output



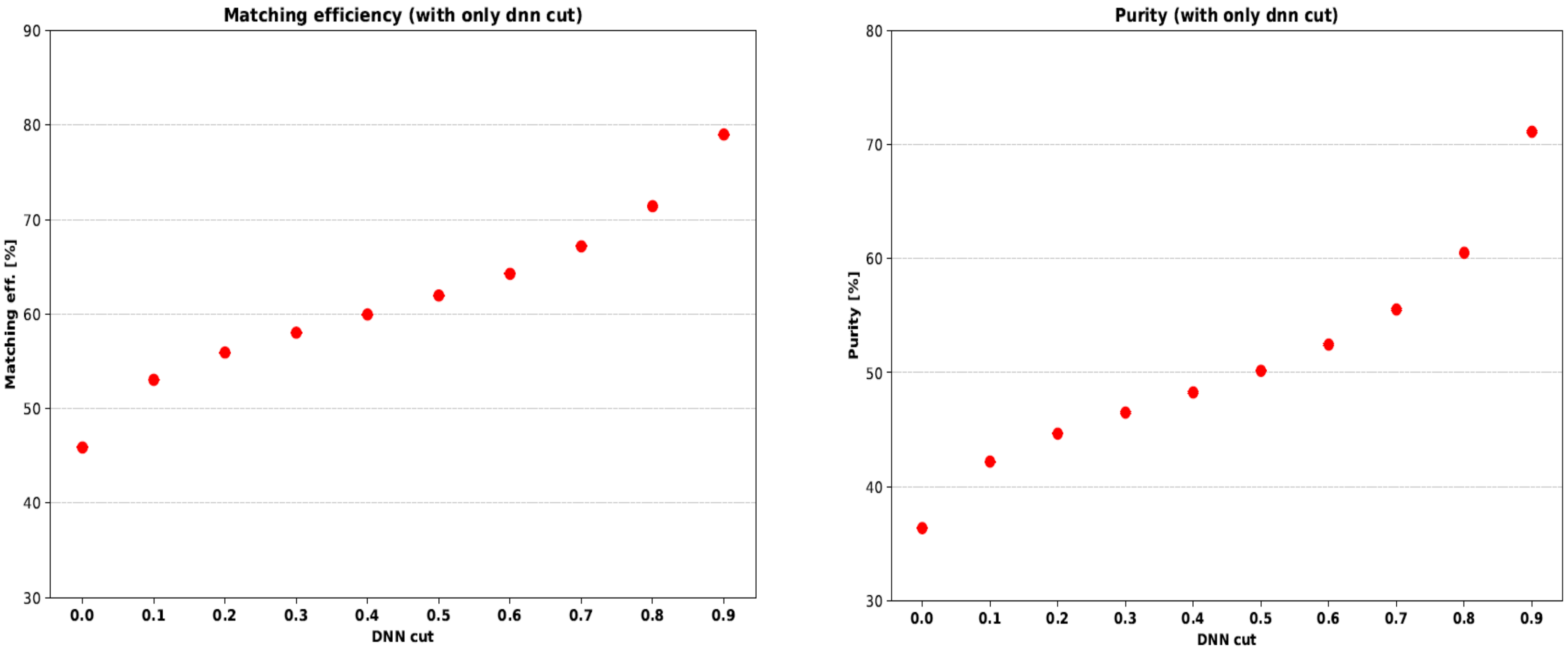
- Sharp peak at $\sim 1 \rightarrow$ DNN confidence in classification.
- Correctly matched (blue) dominate near 1, unmatched (red) at lower DNN scores.
- Can we improve the reconstruction by cutting the on the DNN output?

Impact of DNN cut on reconstruction efficiency



Reconstruction efficiency: Increases from ~79% to ~90% with higher DNN cuts, selecting higher-quality events.

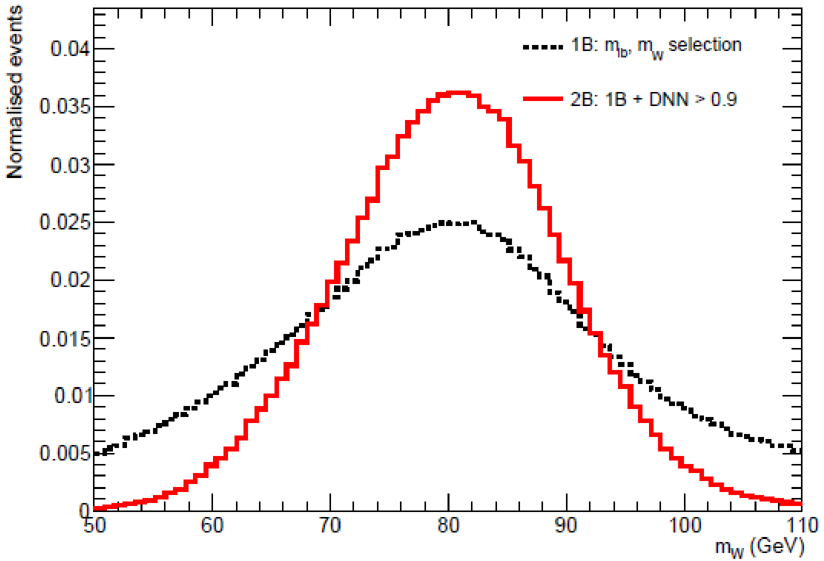
Impact of DNN cut on matching efficiency & purity

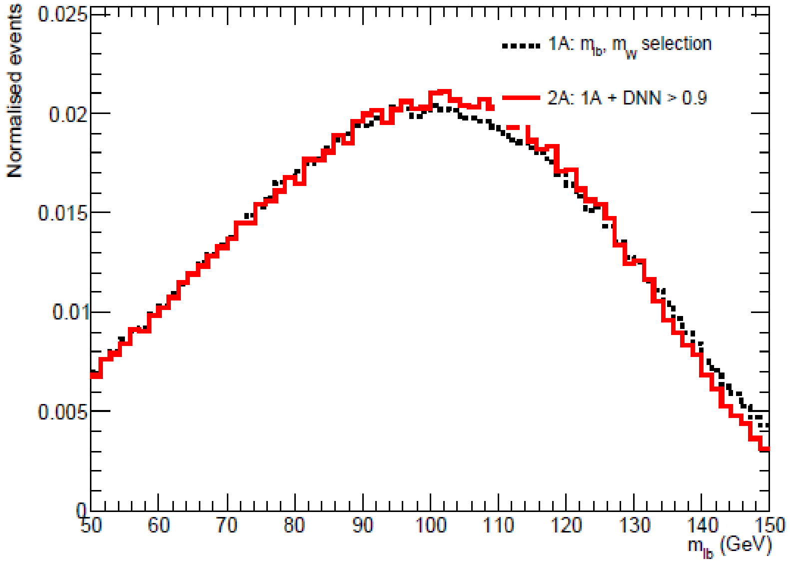


Matching efficiency: Improves with higher DNN cuts, reaching up to ~79%, enhancing signal-to-background ratio.

Purity: Increases to ~71% as DNN cut improves, reducing background contamination.

Effect of DNN cut on m_W and m_{lb} distributions





- DNN cut (less unmatchable events) → m_w distribution becomes more symmetric and narrow.
- **Increased signal-to-background:** Correctly matched events dominate around $m_W \approx 80$GeV.
- Better resolution & sharper mass peak → Improved precision for m_W and m_{lb}.
- For m_{lb} **the effect is small because the efficiency for the b-quark was very high before the cut.**

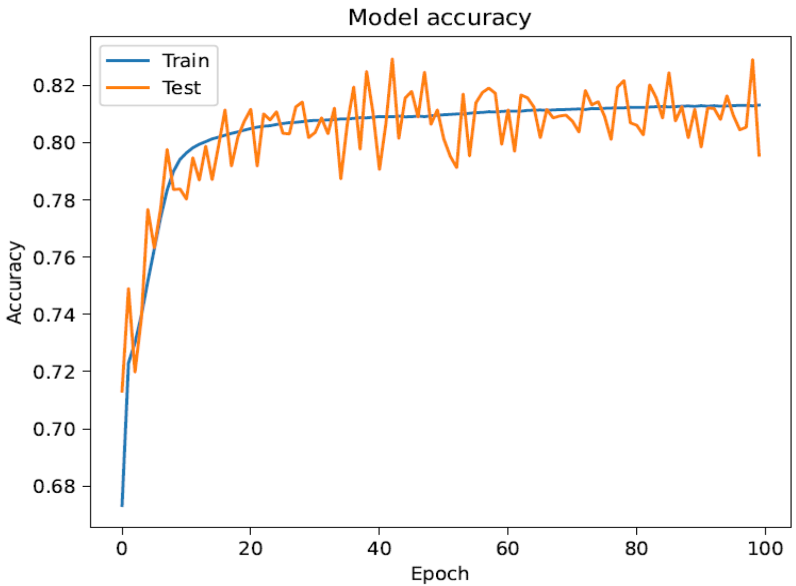
Conclusions

- Reconstruction efficiency, matching efficiency, and purity were enhanced through DNN cut and kinematic selections.
- DNN cuts, combined with m_W and m_{lb} cuts, improved signal-to-background ratio and reduced unmatchable events.
- Clean mass peaks and improved resolution, leading to higher precision in top-quark mass measurement was achieved.
- The study demonstrates that a well-trained DNN provides a robust tool for high-precision top-quark mass measurements at ATLAS.

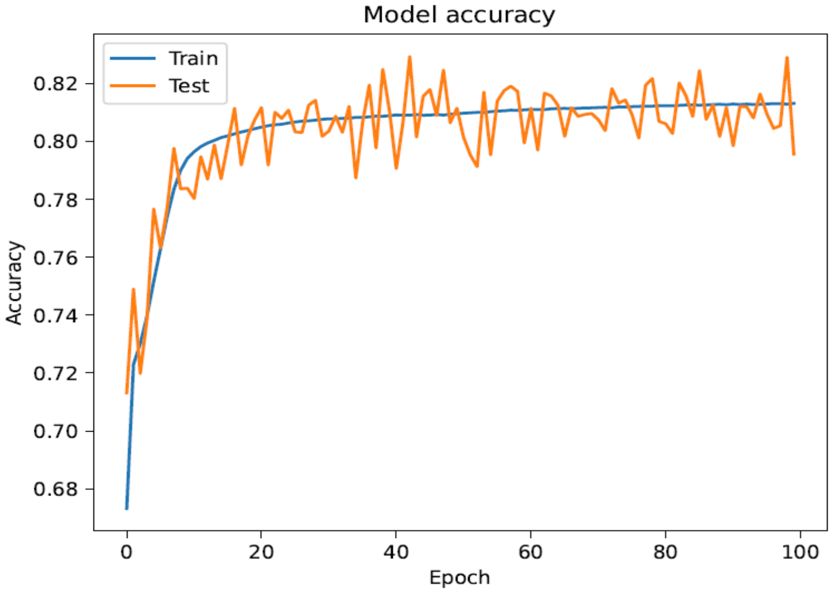
Thank you for listening

BACKUP

Low P_T selection DNN training: Accuracy



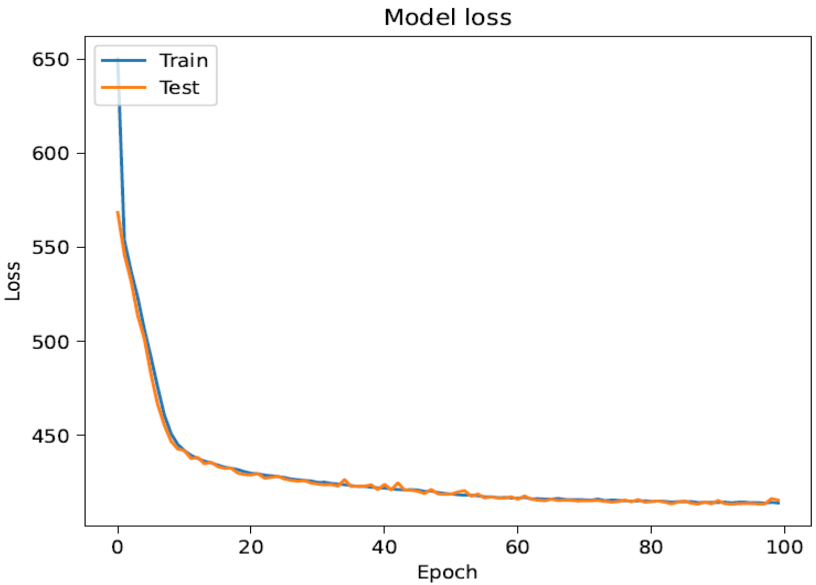
Accuracy for the training on even low P_T selection



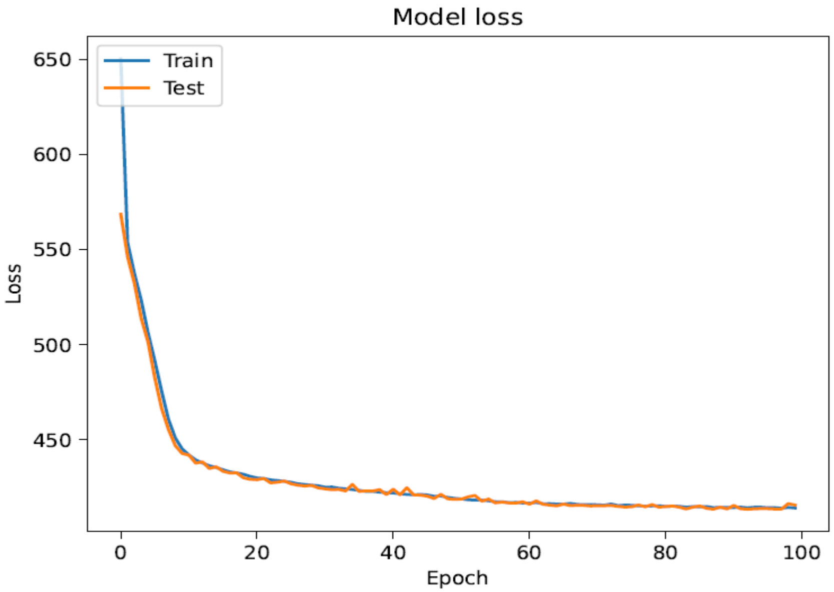
Accuracy for the training on odd low P_T selection

- DNN reaches 80% accuracy after ~30 epochs
- Test accuracy fluctuates but improves, aligning with training.
- Slight **overfitting** observed, with the gap between training and test accuracy narrowing.
- **Convergence** occurs after several epochs, showing stable model learning.

Low P_T selection DNN training: Loss



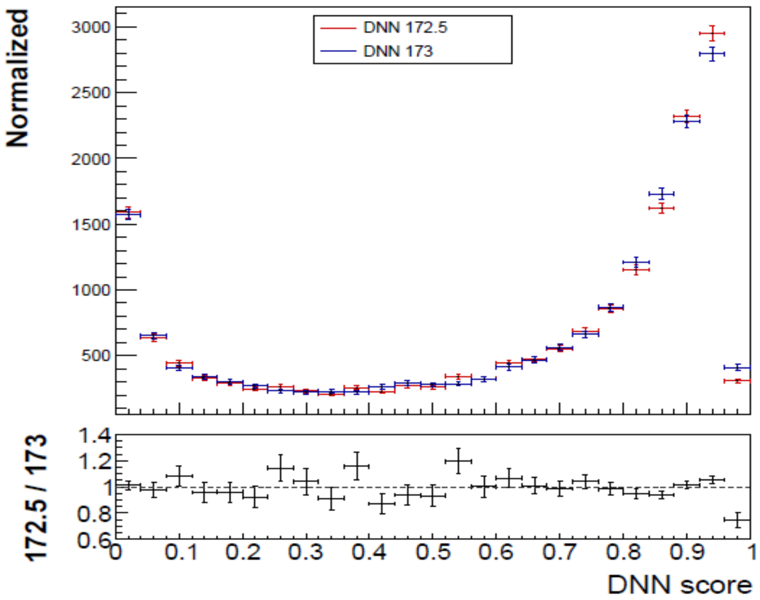
Loss for the training on even low P_T selection

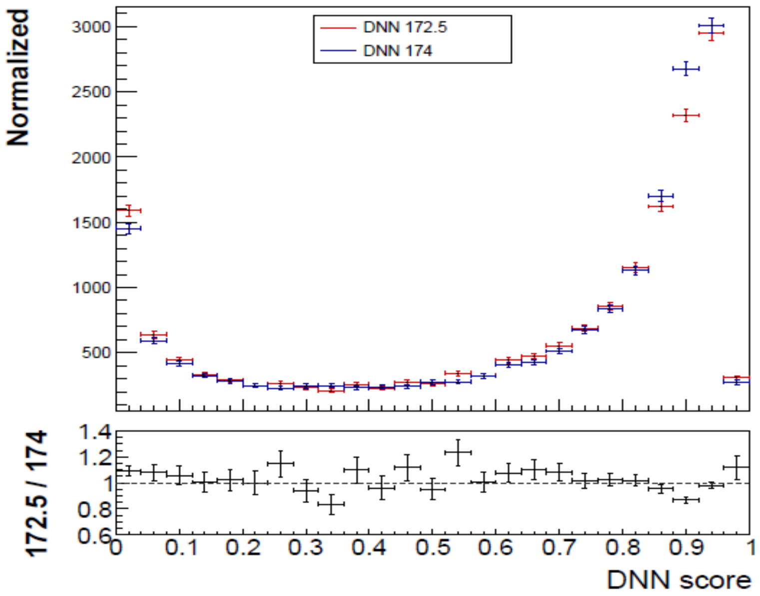


Loss for the training on odd low P_T selection

- Both decreases rapidly from ~600 to ~450.
- Train and test loss are almost identical, indicating good generalization
- Stable convergence, no significant overfitting, with loss flattening as training progresses.
- Model loss stabilizes, showing reliable training and test outcomes.

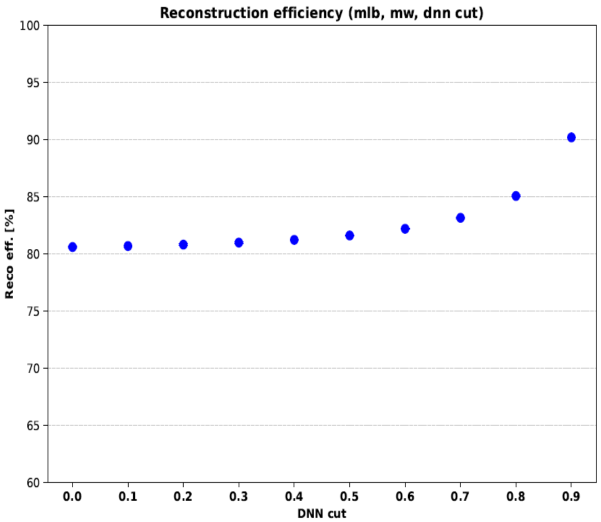
Comparison of m_t hypotheses in the DNN output

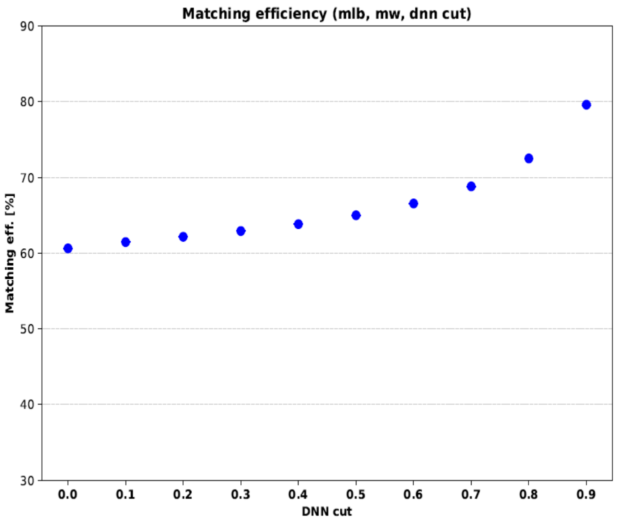


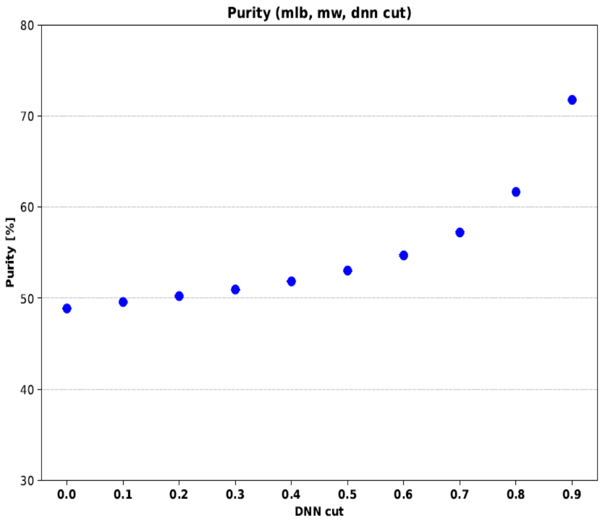


- **DNN score comparison:** Both mass hypotheses show **similar peaks** at high DNN scores.
- **Ratio plots:** Ratios (172.5/173, 172.5/174) stay **near 1**, indicating **minimal mass difference**.
- **DNN confidence:** The DNN **confidently classifies** the masses with **minimal variation**.

Effect of m_W, m_{lb}, & DNN cuts on efficiency and purity

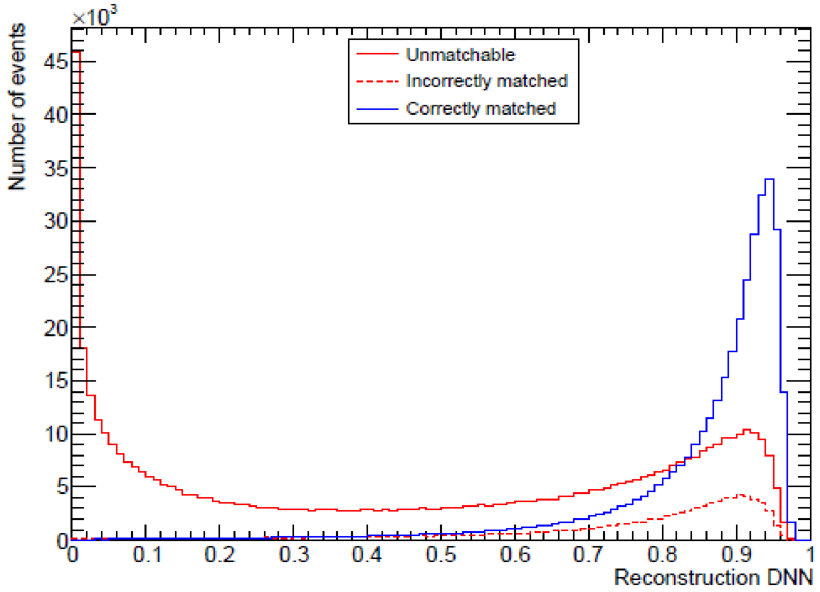




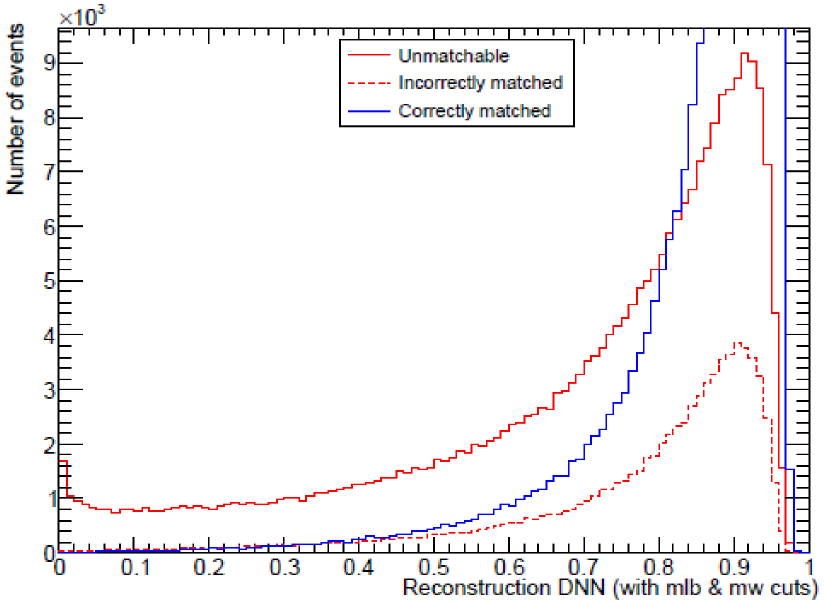


- **Reconstruction efficiency:** Steady increase from ~80% to ~90% as DNN, m_W, and m_{lb} cuts are applied, selecting higher-quality events.
- **Matching efficiency:** Increases to ~80%, with higher cuts improving the signal-to-background ratio by removing incorrect matches.
- **Purity:** Improves to ~72%, indicating a significant reduction in background contamination as the DNN cut increases.

Reconstruction DNN Output



- Sharp peak at $\sim 1 \rightarrow$ DNN confidence in classification.
- Correctly matched (blue) dominate near 1, unmatched (red) at lower DNN scores.
- About **85%** of events correctly matched near 1.



- Sharper peak for correctly matched events.
- Reduced incorrectly matched/unmatched events after cuts.
- Effective kinematic cuts improve purity without significant efficiency loss.