GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Classification of signal and background in associated $t\bar{t}H$ production via a multi-class neural network.

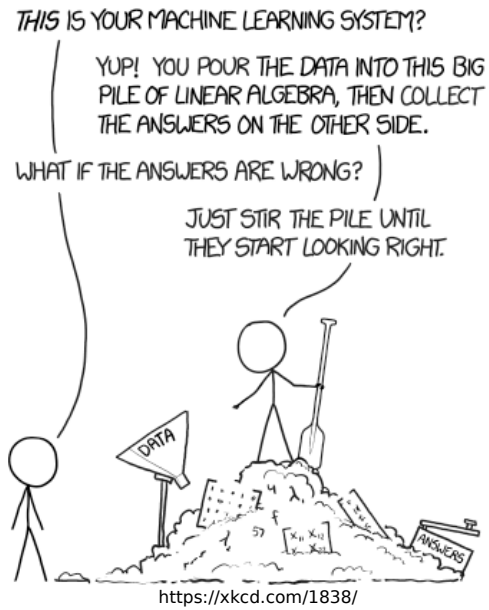
Konrad Helms

University of Göttingen

Supervised by Prof. Dr. A. Quadt

July 16, 2021

BSc Student Seminar Talk, 2021



https://xkcd.com/1838/

Contents

Introduction

Introduction

- mid-20th century: Standard Model
- rapid technological advances:

  $\rightarrow$ 1995: top quark t discovered by DØ and CDF

  $\rightarrow$ 2012: Higgs boson H discovered by ATLAS and CMS
- since then, many measurements of the Higgs boson's properties to find constraints on couplings

  $\rightarrow$ 2018: observation of Higgs-associated top anti-top quark pair production $t\bar{t}H$

Production Modes and Background Processes

Production Modes and Background Processes

- top-Higgs coupling not only sensitive to properties of the Higgs boson, but also to yet undiscovered particles
- in the past for example: $H \to b\bar{b}$, $H \to W^+W^-$
- not possible with the top quark, $m_H \approx 125\,\text{GeV}$, $m_t \approx 173\,\text{GeV}$
- instead of the decay, we measure $t\bar{t}H$ production (and tH production)

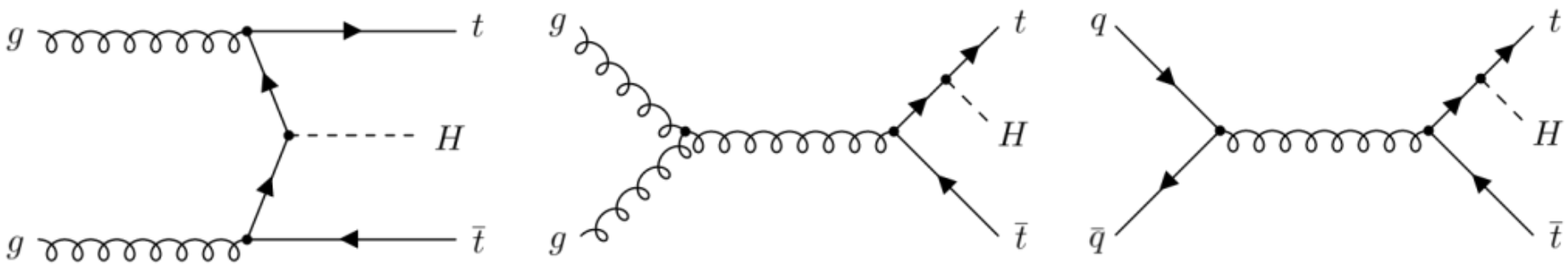


Figure: Three possible tree level Feynman diagrams in proton-proton collisions for the $t\bar{t}H$ production mode.

Production Modes and Background Processes

- another way of measuring the top-Higgs coupling: Higgs boson production in association with a single top quark tH

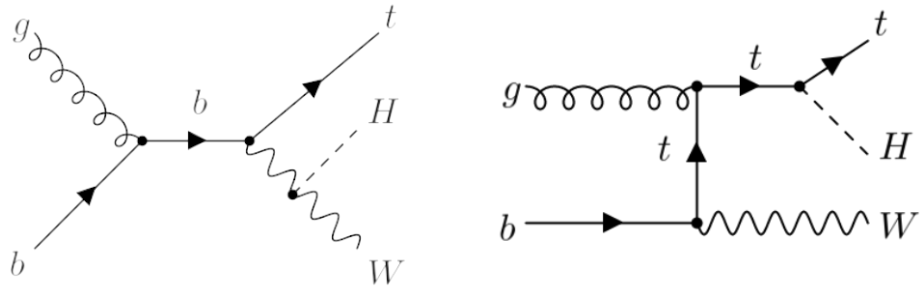


Figure: Representative Feynman diagrams for the tH production with an on-shell W boson in the final state.

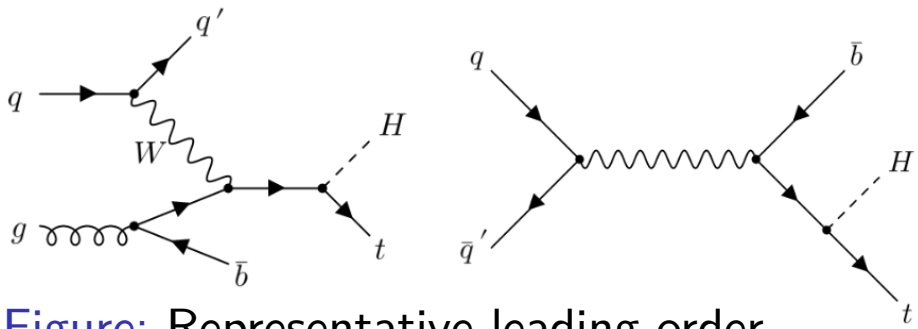


Figure: Representative leading order Feynman diagrams for the tH production via the t- and s-channel.

Production Modes and Background Processes

- most prominent background processes: $t\bar{t} + b$, $t\bar{t} +$ light jets, $t\bar{t} + c$

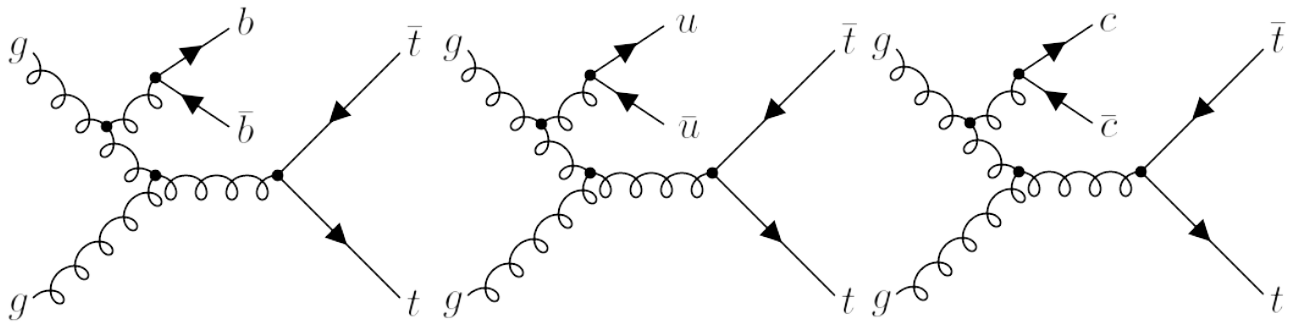


Figure: Representative Feynman diagrams for the $t\bar{t} + b$, $t\bar{t} +$ light jets, $t\bar{t} + c$.

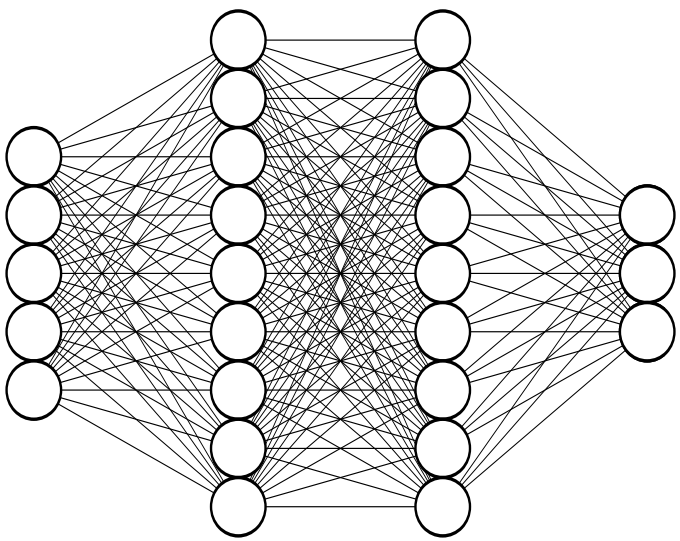
- top quark decay dominated by $t \to Wb$ with subsequent W boson decays: $W \to q\bar{q}'$ or $W \to \ell\nu$

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Neural Networks

- nodes hold an activation a, connections are assigned a weight w and a bias b
- weighted sum: $\sum_i a_i \cdot w_i + b$
- apply activation functions Sigmoid (1d), Softmax (multi-dim.), ReLU (1d)

  $\rightarrow f\left(\sum_i a_i \cdot w_i + b\right)$



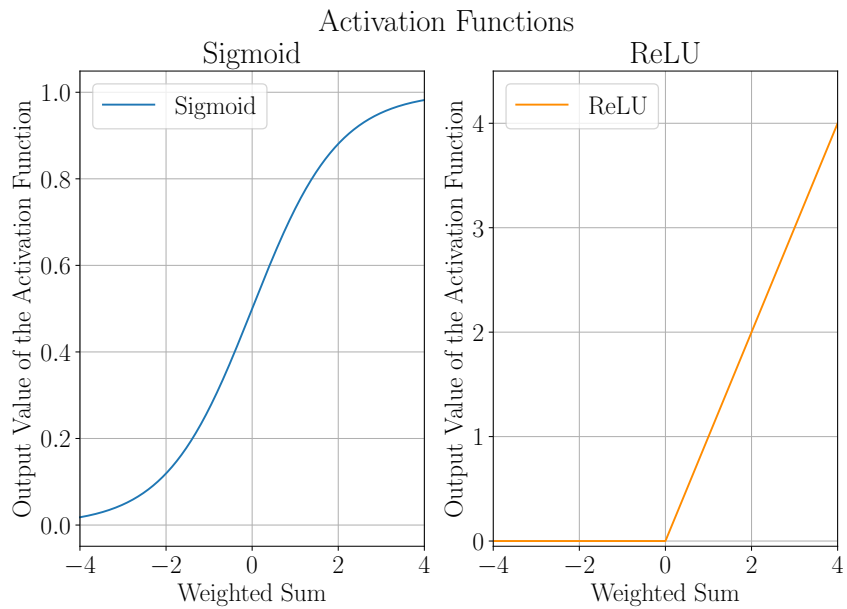


Figure: Neural network structure (left) with a 5 node input layer, 2 hidden layers, 3 node output layer and two activation functions (right).

- cost function C: input = all weights and biases, parameterised by neural network's behaviour over a number of training events, output = cost
- for a single training example: loss $= -\sum_{i=1}^{K} y_i \cdot \ln(a_i)$, a_i activations, y_i correct activation of i-th output node
- learning = minimising the cost function C, in this thesis $\sim 10^3$ inputs
- gradient descent: calculate $-\nabla C$, take a small step, repeat until we reach a minimum

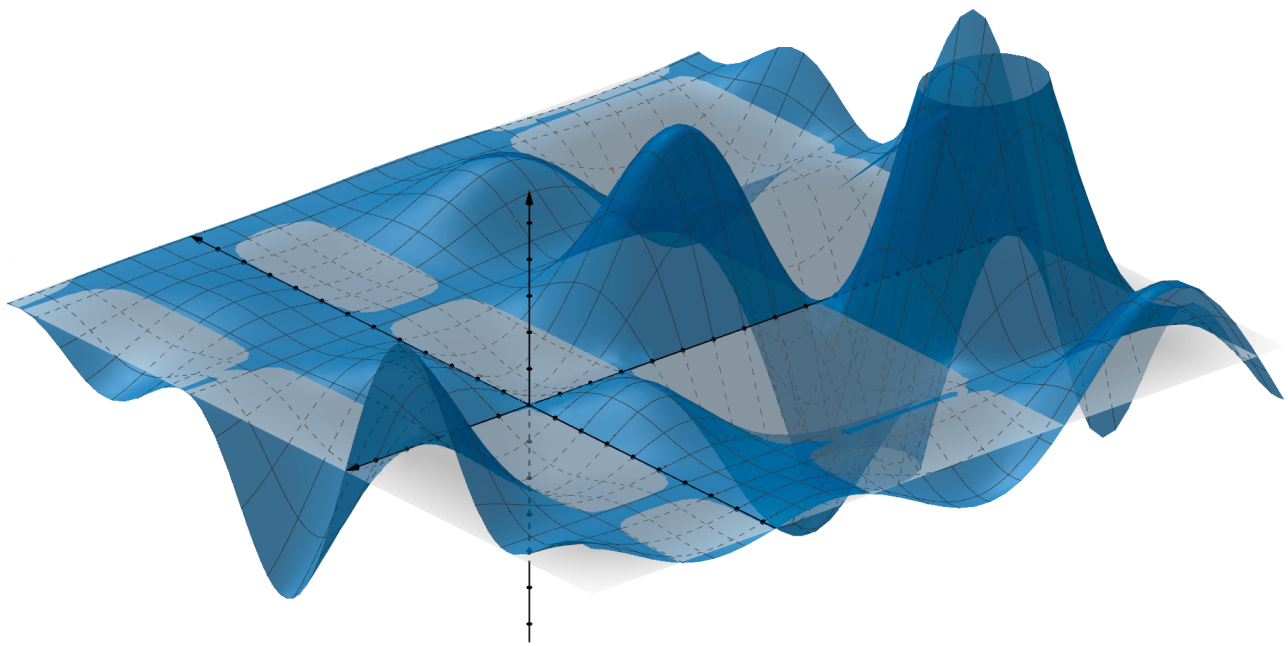


Figure: Representative cost function having 2 inputs.

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Results of the Bachelor's Thesis

Results of the Bachelor's Thesis

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

- 2 datasets:
  - small dataset: up to 280,000 Monte Carlo events
  - large dataset: up to 1,900,000 Monte Carlo events

- small dataset:

Region	#leptons	#jets	#b-tags @60%	@70%	@77%	@85%
6 jet, high $^{\geq 6j}_{\geq 4b}$	$=1$	≥ 6	≥ 4	≥ 4	≥ 4	≥ 4
6 jet, low $^{\geq 6j}_{< 4b}$			< 4			
5 jet, high $^{5j}_{\geq 4b}$		$=5$	≥ 4			
5 jet, low $^{5j}_{< 4b}$			< 4			

Table: The non-overlapping event regions of the small dataset.

- large dataset:

Region	#leptons	#jets	# b-tags at 85% WP
6 jet $^{\geq 6j}_{\geq 4b}$	$=1$	≥ 6	≥ 4
5 jet $^{\geq 6j}_{\geq 4b}$		$= 5$	≥ 3

Table: The non-overlapping event regions of the large dataset.

- binary class. neural networks: $t\bar{t}H$ vs. background (excluding tH)
- 5-class neural networks: $t\bar{t}H$, tH, $t\bar{t}+b$, $t\bar{t}+c$, $t\bar{t}+$ light jets
- 4-class neural networks: $t\bar{t}H$, tH, $t\bar{t}+b$, $t\bar{t}+$ light jets and c
- 3-class neural networks: $t\bar{t}H$, tH, $t\bar{t}+$ jets

Results of the Bachelor's Thesis

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Results of the binary class. neural networks:

- small dataset:

Neural network (NN)	Performance increase [%]		
	avg. training	avg. testing	avg. overall
3 layer NN, 6 jet incl.	0.53	0.45	0.49
5 layer NN, 6 jet incl.	0.95	0.71	0.83
3 layer NN, 6 jet high	1.08	0.97	1.03

Table: Average performance increase with respect to the BDTs used in previous analyses.

- large dataset:
  - performance decrease by 0.26%
  - but: BDTs trained on older ntuples and tested on tighter region
  - $\Longrightarrow$ only partially comparable results!

Results of the Bachelor's Thesis

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Results of the multi-class neural networks:

- small dataset:

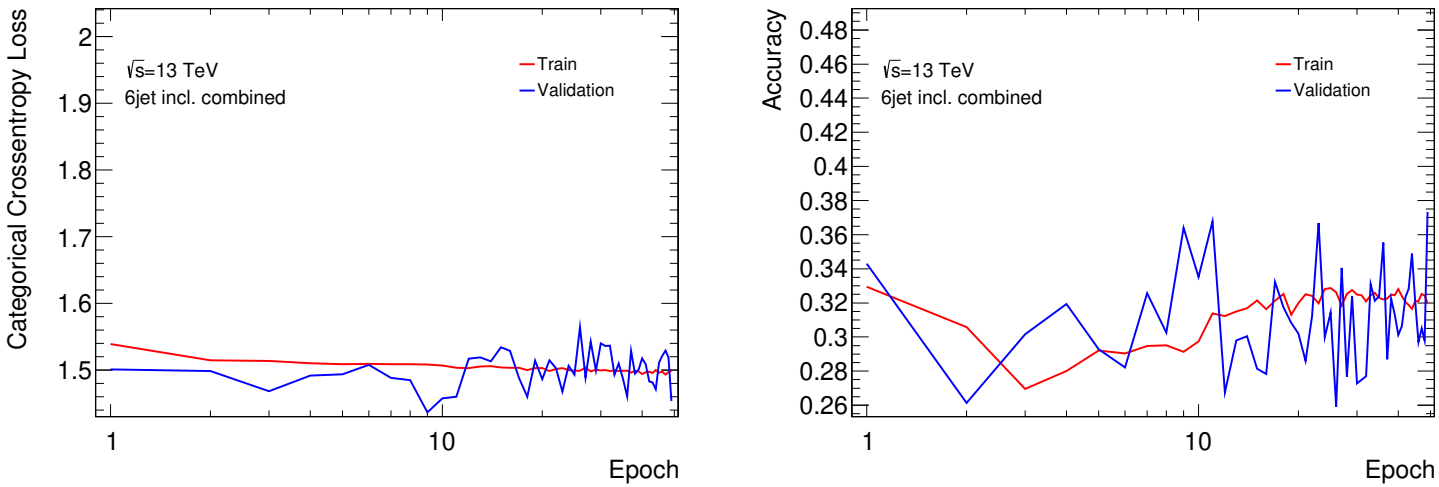


Figure: Representative categorical crossentropy loss function (left) and accuracy (right) for the 5-class neural network.

Results of the Bachelor's Thesis

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Results of the multi-class neural networks:

- large dataset: (3-class neural network delivered the best results)

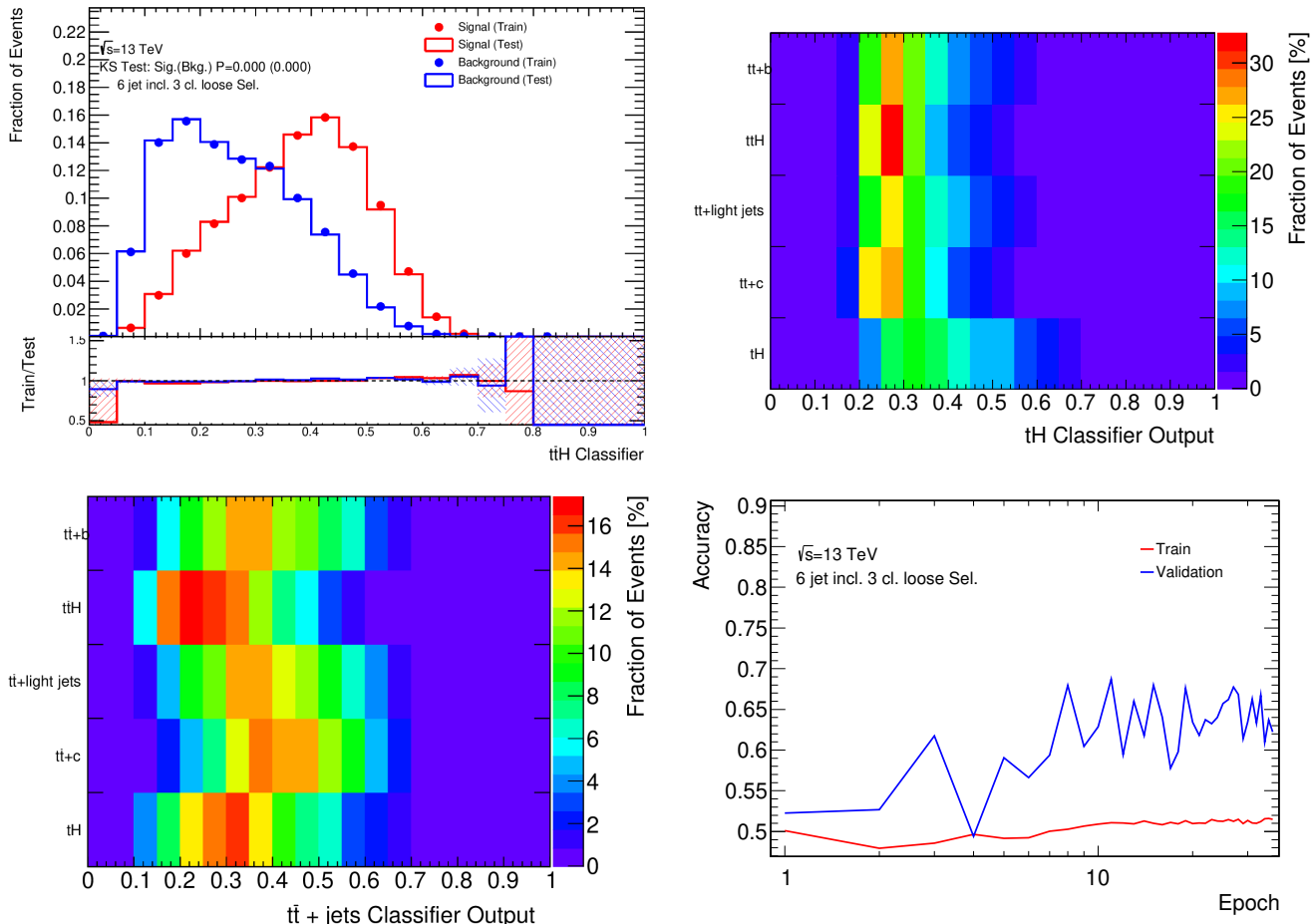


Figure: Representative $t\bar{t}H$ Train/Test plot (top left), tH confusion plot (top right), $t\bar{t}$ + jets confusion plot (bottom left) and accuracy plot (bottom right).

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Thank you for your attention.

Appendix

Monte Carlo Statistics

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Class	Number of Monte Carlo events					
	5 jet low	5 jet high	6 jet low	6 jet high	5 jet	6 jet
	smaller dataset				larger dataset	
$t\bar{t}H$	81,160	85,081	205,230	215,050	1,407,148	1,236,630
tH	3,270	2,853	7,336	6,978	83,478	47,006
$t\bar{t}$ + light jets	10,284	7254	27557	22063	1,600,000	3,607,537
$t\bar{t} + b$	136,987	97,837	278,391	235,345	1,600,000	1,919,591
$t\bar{t} + c$	3,204	382	6410	771	1,600,000	1,324,606

Table: Number of Monte Carlo events in each class and region in the smaller and larger dataset.

Event Yields

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Region	Event yield ± Poisson error				
	tH	$t\bar{t}H$	$t\bar{t}$ + light jets	$t\bar{t}+c$	$t\bar{t}+b$
Small dataset					
6 jet, high$^{\geq 6j}_{\geq 4b,\text{hi}}$	3.4 ± 1.8	213.1 ± 14.6	$2,638.7 \pm 51.4$	109.8 ± 10.5	$2,669.4 \pm 51.7$
6 jet, low$^{\geq 6j}_{< 4b,\text{lo}}$	3.3 ± 1.8	207.0 ± 14.4	$3,239.6 \pm 56.9$	765.4 ± 27.7	$3,131.3 \pm 56.0$
5 jet, high$^{5j}_{\geq 4b,\text{hi}}$	3.4 ± 1.8	64.4 ± 8.0	897.7 ± 30.0	55.0 ± 7.4	$1,013.6 \pm 31.8$
5 jet, low$^{5j}_{< 4b,\text{lo}}$	3.3 ± 1.8	64.4 ± 8.0	$1,240.7 \pm 35.2$	396.8 ± 19.9	$1,403.9 \pm 37.5$
6 jet, high & low combined	6.7 ± 2.6	420.1 ± 20.5	$5,878.3 \pm 76.7$	875.2 ± 29.6	$5,800.7 \pm 76.2$
Large dataset					
6 jet$^{\geq 6j}_{\geq 4b,@85\%}$	22.3 ± 4.7	$1,311.7 \pm 36.2$	$50,897.0 \pm 225.6$	$30,689.8 \pm 175.2$	$22,354.3 \pm 149.5$
5 jet$^{\geq 6j}_{\geq 3b,@85\%}$	81.4 ± 9.0	$1,277.6 \pm 35.7$	$108,241.0 \pm 329.0$	$125,979.4 \pm 354.9$	$54,153.2 \pm 231.7$

Table: The event yield and Poisson error in each category of the small and large datasets.

Further Plots:
Binary Classification Neural Networks

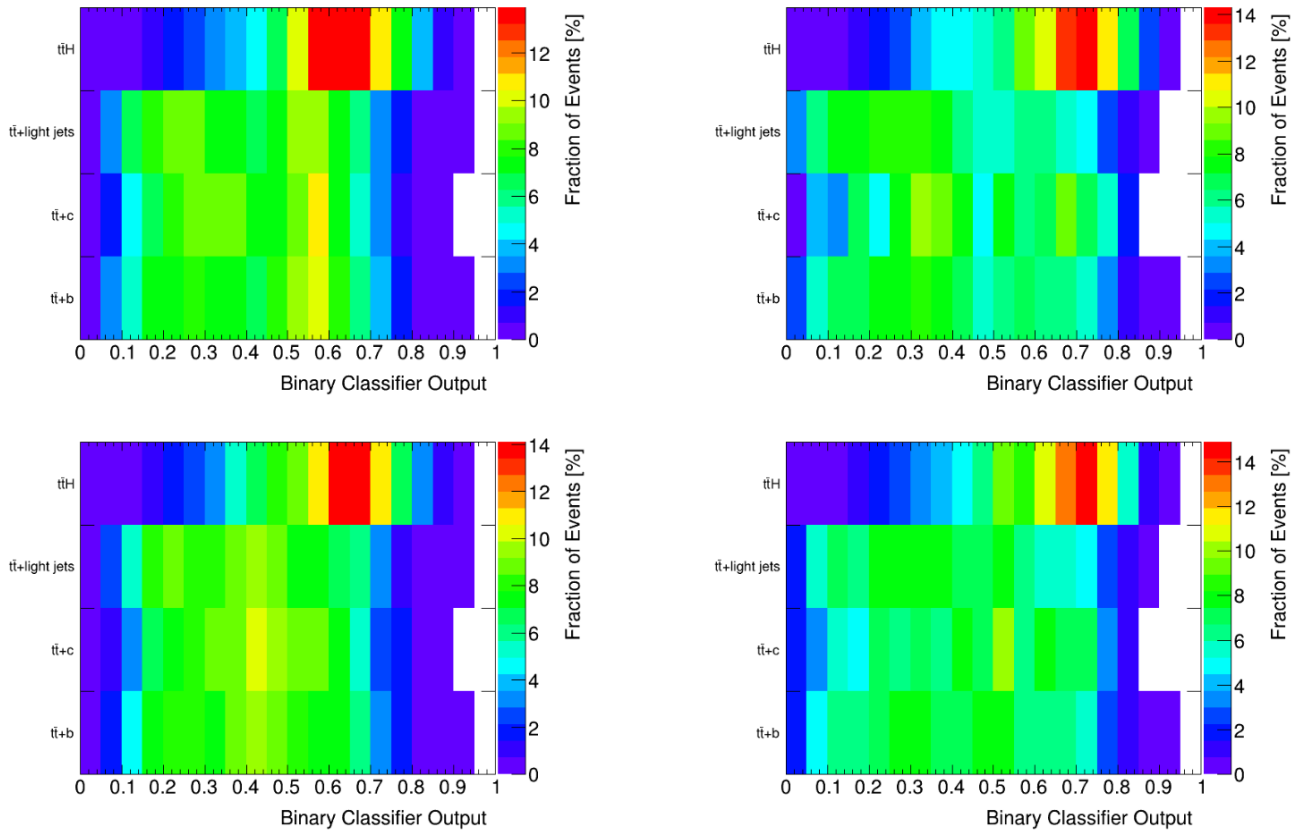


Figure: Confusion plots of the binary classification neural networks, trained on the 5 jet low (top left) and high (top right) region, and the 6 jet inclusive low (bottom left) and high (bottom right) region of the small dataset.

Further Plots:
Binary Classification Neural Networks

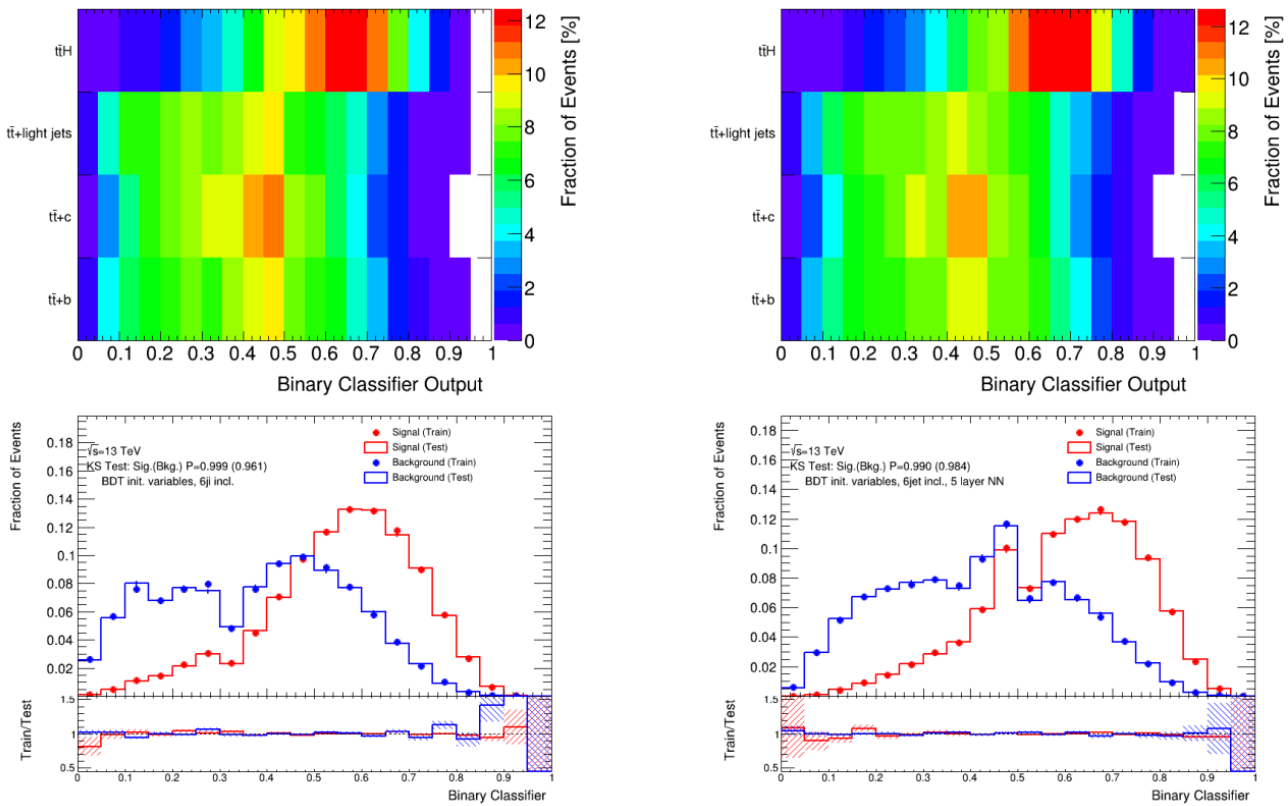
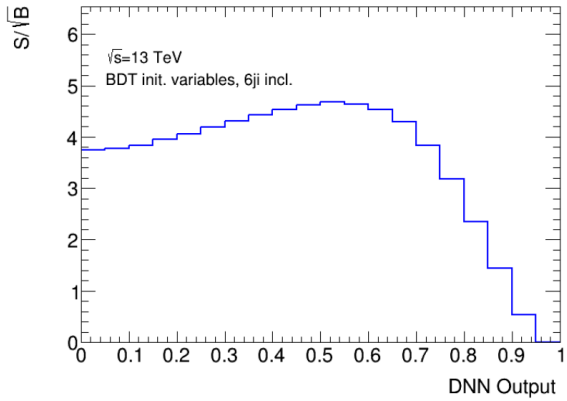


Figure: Confusion plots (top) and representative Train/Test plots (bottom) of the 3 layer (left) and 5 layer (right) binary classification neural networks, trained on the 6 jet inclusive combined region.

Further Plots:
Binary Classification Neural Networks



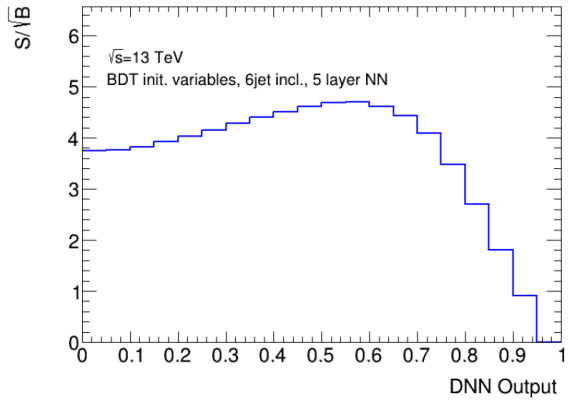


Figure: Representative signal-over-square-root-background plots for the 3 layer (left) and 5 layer (right) binary classification neural networks, trained on the 6 jet inclusive combined region of the small dataset.

Further Plots:
Binary Classification Neural Networks

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

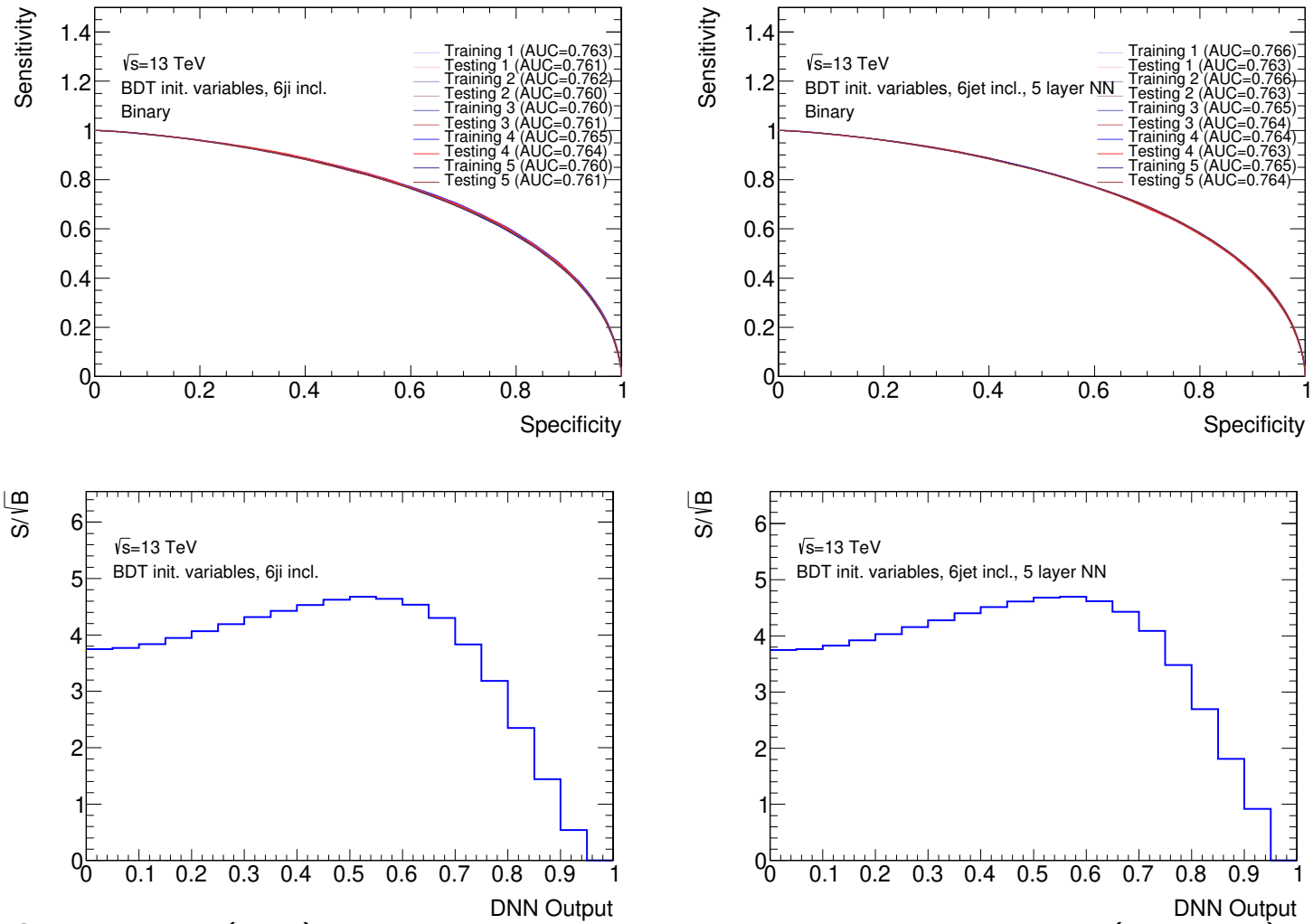


Figure: ROC curves (top) and signal-over-square-root-background (bottom) plots of the 3 layer (left) and 5 layer (right) binary class. neural networks, trained on the small dataset.

Further Plots:

5-class Neural Network, small dataset

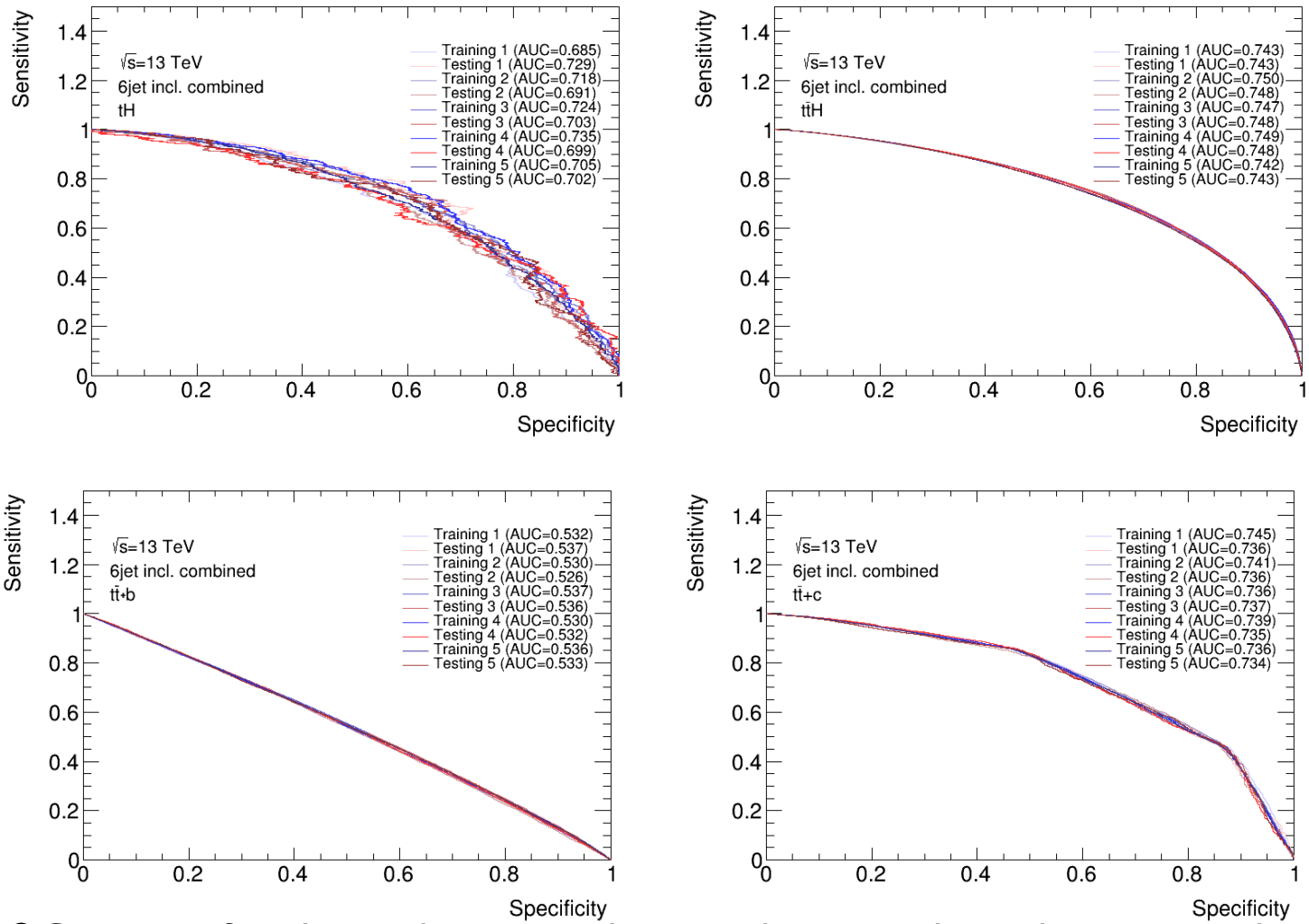


Figure: ROC curves for the 5-class neural network, trained on the 6 jet inclusive combined region of the small dataset.

Further Plots:

5-class Neural Network, small dataset

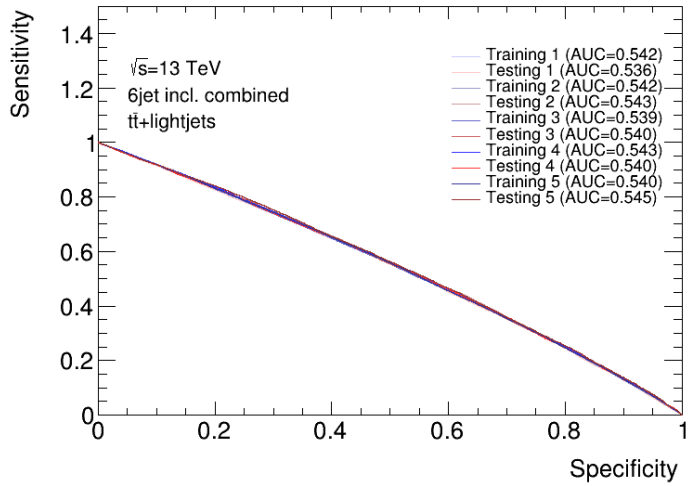


Figure: ROC curve for the 5-class neural network, trained on the 6 jet inclusive combined region of the small dataset.

Further Plots:

4-class Neural Networks

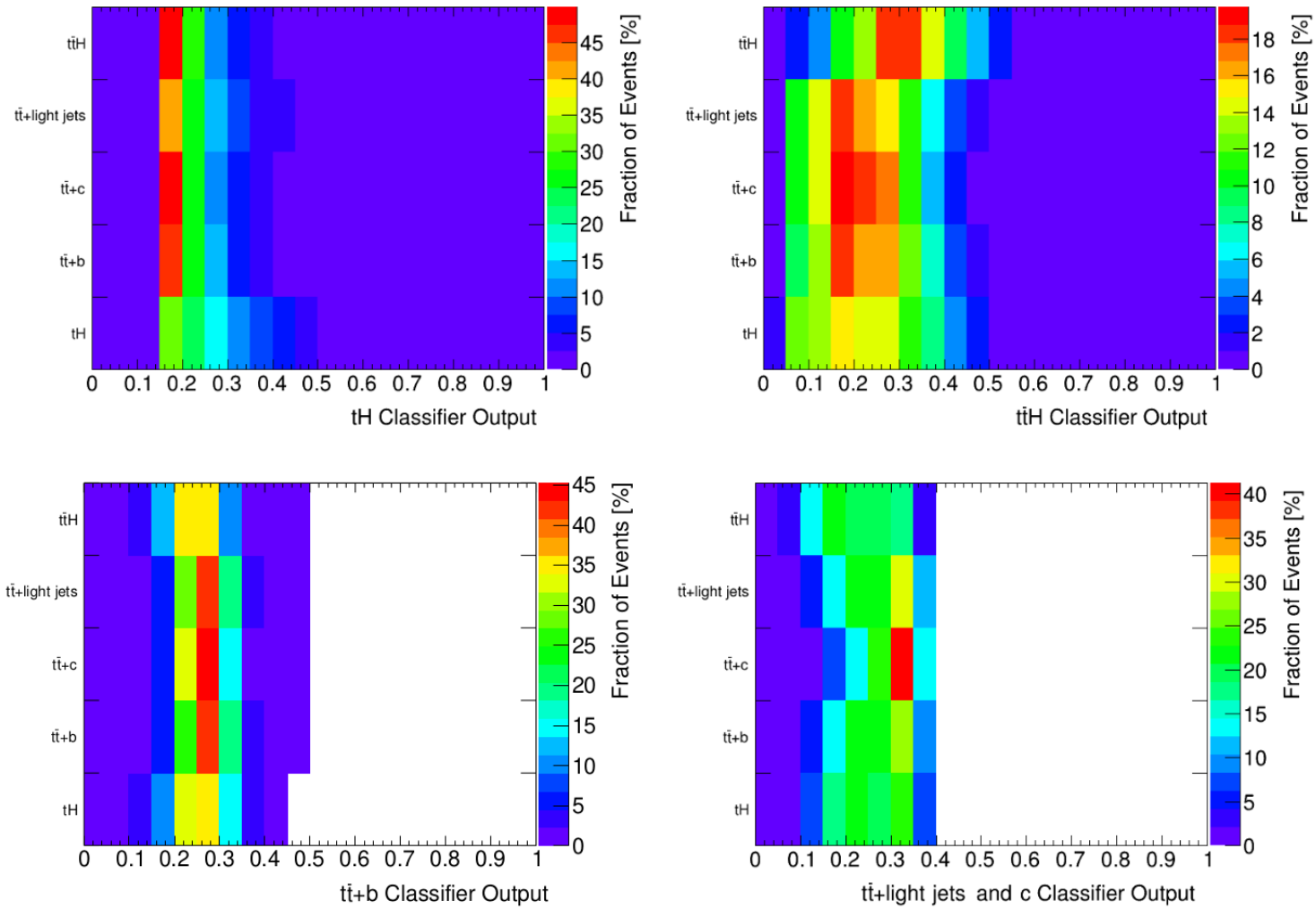


Figure: Confusion plots for the 4-class neural network, trained on the 5 jet region of the large dataset.

Further Plots:

4-class Neural Networks

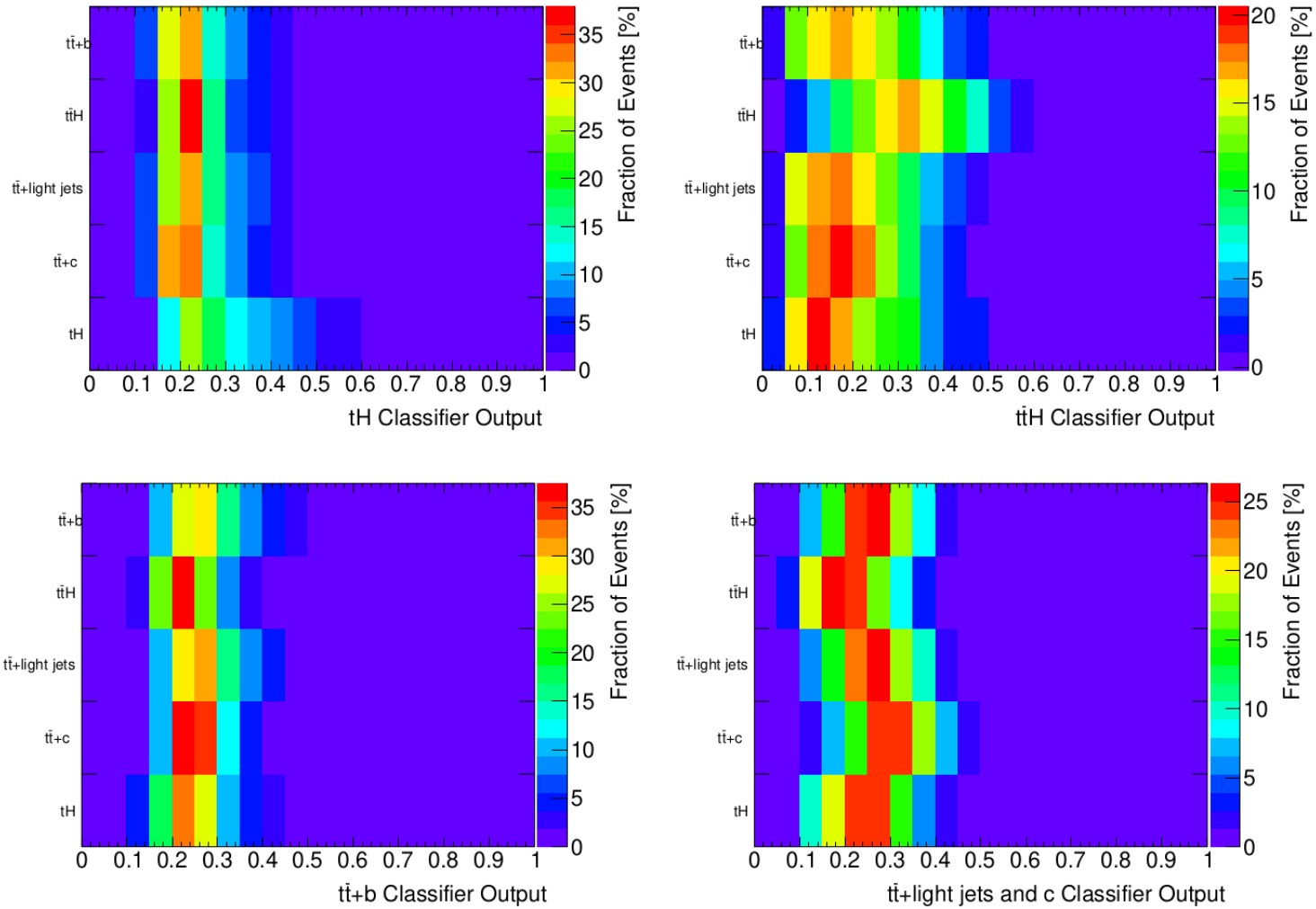


Figure: Confusion plots for the 4-class neural network, trained on the 6 jet inclusive region of the large dataset.

Further Plots:

4-class Neural Networks

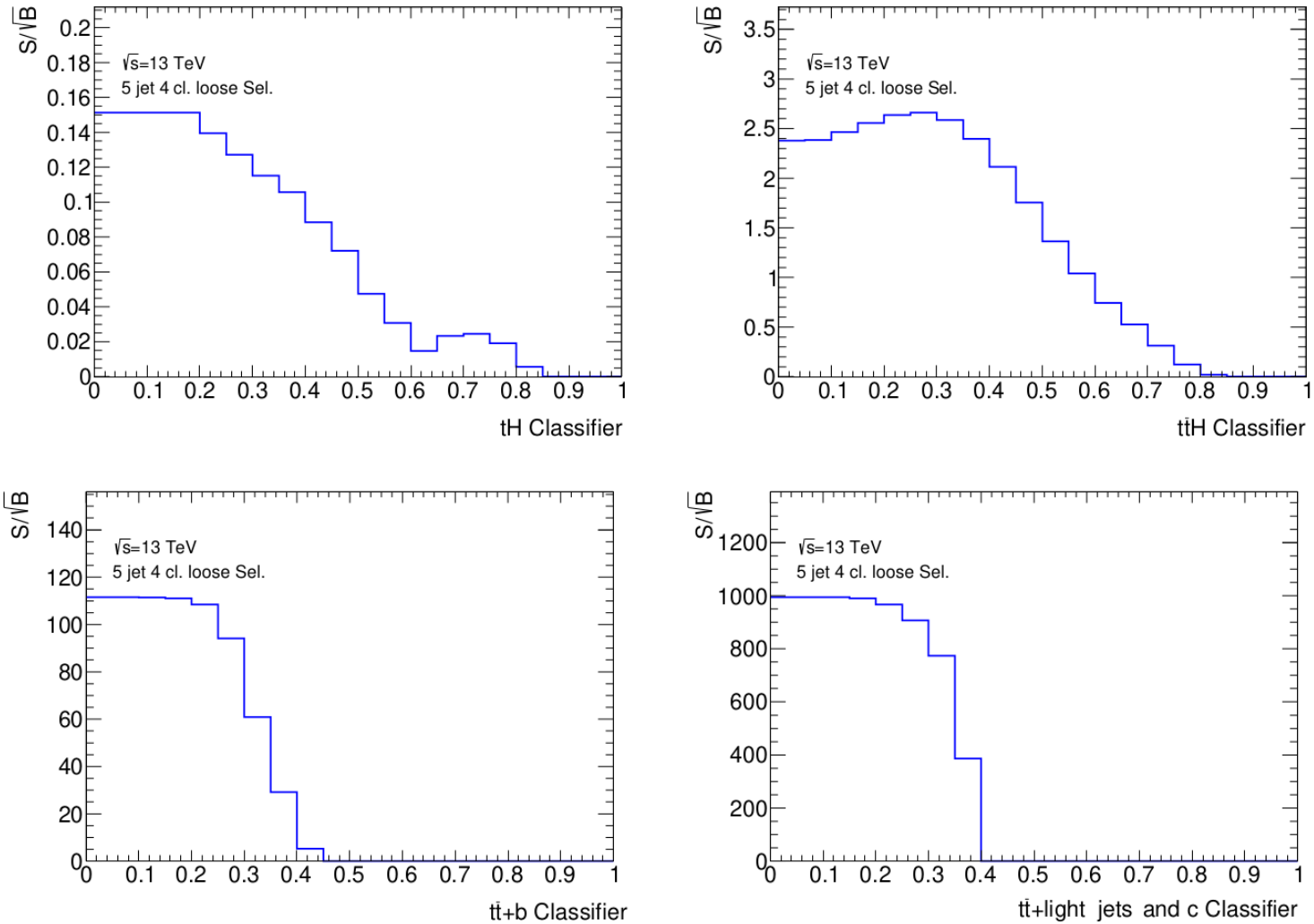


Figure: Signal-over-square-root-background plots for the 4-class neural network, trained on the 5 jet region of the large dataset.

Further Plots:

4-class Neural Networks

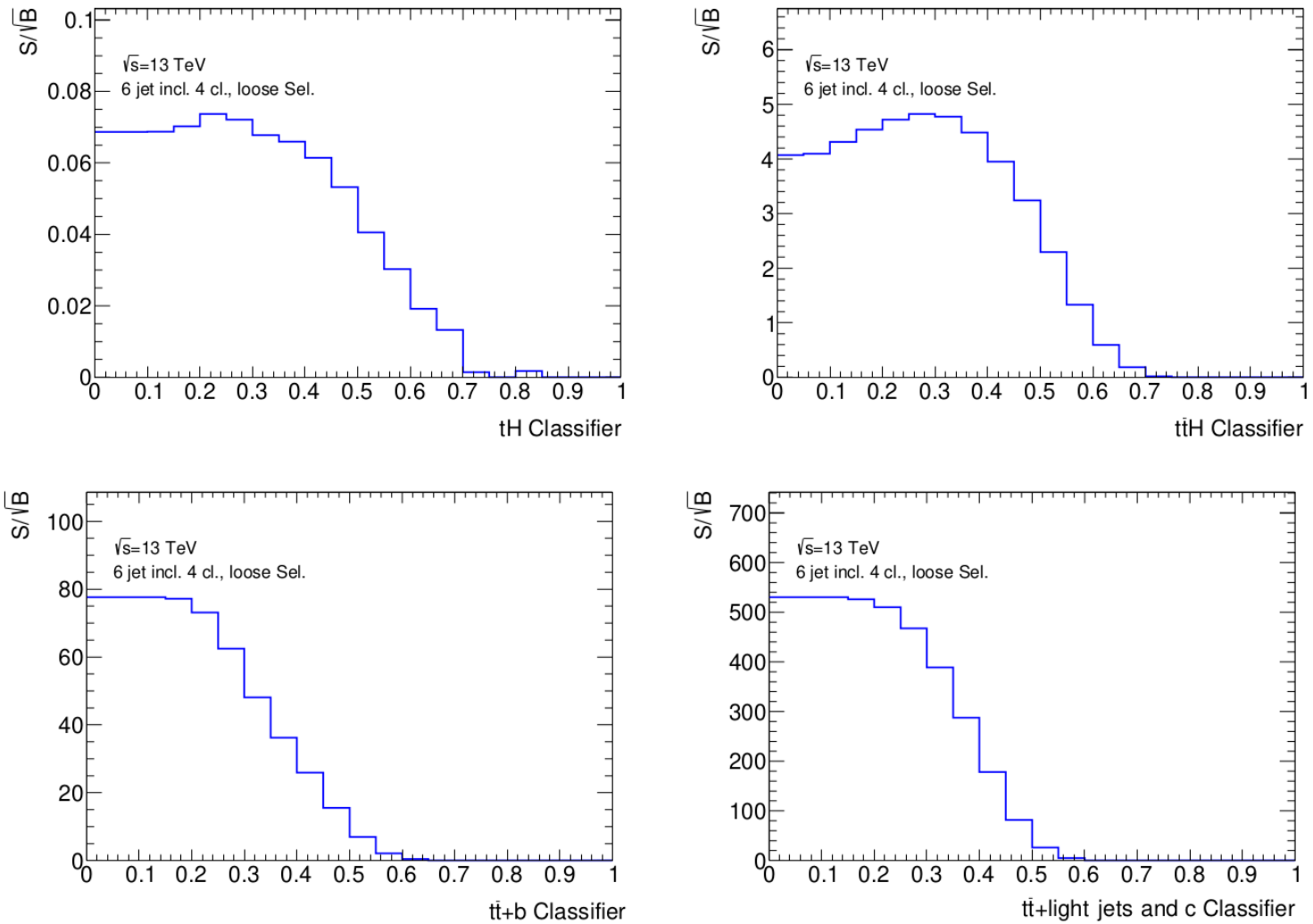


Figure: Signal-over-square-root-background plots for the 4-class neural network, trained on the 6 jet inclusive region of the large dataset.

Further Tables:

3-class Neural Networks

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

	5 jet			6 jet inclusive		
	AUC avg. training	AUC avg. testing	AUC avg. overall	AUC avg. training	AUC avg. testing	AUC avg. overall
$t\bar{t}H$	0.7095	0.7090	0.7093	0.7400	0.7400	0.7400
tH	0.5918	0.5888	0.5903	0.6938	0.6898	0.6918
$t\bar{t}+\text{jets}$	0.6875	0.6875	0.6875	0.7260	0.7259	0.7260

Table: Average AUCs of the ROC curves during training, testing and overall average of the 3-class NN, trained on the large dataset, for each output class.

Further Tables:

4-class Neural Networks

	5 jet			6 jet inclusive		
	AUC avg. training	AUC avg. testing	AUC avg. overall	AUC avg. training	AUC avg. testing	AUC avg. overall
$t\bar{t}H$	0.7193	0.7193	0.7193	0.7410	0.7163	0.7287
tH	0.6135	0.6100	0.6118	0.7020	0.6963	0.6992
$t\bar{t}+b$	0.5313	0.5313	0.5313	0.5318	0.5318	0.5318
$t\bar{t}+$ light and c	0.5695	0.5693	0.5694	0.5833	0.5833	0.5833

Table: Average AUCs of the ROC curves during training, testing and overall average of the 4-class NN, trained on the large dataset, for each output class.

Further Tables:

5-class Neural Networks

	5 jet			6 jet inclusive		
	AUC avg. training	AUC avg. testing	AUC avg. overall	AUC avg. training	AUC avg. testing	AUC avg. overall
$t\bar{t}H$	0.7235	0.7230	0.7233	0.7483	0.7480	0.7482
tH	0.6295	0.6265	0.6280	0.6953	0.6928	0.6941
$t\bar{t}+b$	0.5478	0.5480	0.5479	0.5778	0.5778	0.5778
$t\bar{t}+c$	0.6333	0.6333	0.6333	0.7393	0.7388	0.7391
$t\bar{t}+$ light jets	0.5740	0.5738	0.5739	0.5948	0.5950	0.5949

Table: Average AUCs of the ROC curves during training, testing and overall average of the 5-class NN, trained on the large dataset, for each output class.