technische universität
dortmund

Arbeit zur Erlangung des akademischen Grades
Bachelor of Science

Neural network based signal-background classification for the differnetial single top+photon measurement at the ATLAS experiment

Cihad Gözsüz
geboren in Dortmund

2021

Lehrstuhl für Experimentelle Physik IV
Fakultät Physik
Technische Universität Dortmund

Erstgutachter:	Prof. Dr. Johannes Erdmann
Zweitgutachter:	Prof. Dr. Johannes Albrecht
Abgabedatum:	13. July 2021

Abstract

The abstract is a short summary of the thesis in English, together with the German summary it has to fit on this page.

Contents

1 Introduction

The Standard Model (SM) of particle physics describes the nature of discovered elementary particles and three of the four fundamental interactions: the strong, weak and electromagnetic interaction. Only the gravitational interaction is not covered by the theory. The SM has been researched extensively and is widely regarded as the most successful theory. However, the SM contains several conceptional problems and is unable to explain all phenomena. For instance, the existence of dark matter is not accounted for in the SM. The many problems of the SM motivate the search for physics beyond the Standard Model (BSM). It is therefore essential to test and research the limits of this theory.
Many tests of the Standard Model involve the top quark, the most massive elementary particle. One such test would be the search for the single production of a top quark in association with a photon in proton-proton collisions ($pp \rightarrow tq\gamma$). The process is sensitive to the electroweak coupling of the top quark. As this coupling is a critical parameter of the SM, a precise measurement of the cross-section for this process may give new insights into this parameter. The $tq\gamma$ process has not yet been observed but the CMS Collaboration reported evidence for it in 2018 corresponding to 4.4 standard diviations [14]. Since the discovery of the process should be possible with the full Run-2 dataset, studies with regard to a differential measurement are carried out in this thesis. The differnetial measurement would yield a close examination of the structure of the electromagnetic coupling of the top quark.
As $tq\gamma$ is a rare process of the SM, a significant amount of background occupies the measurement region and the signal to background ratio is inherently small. A classifying neural network (NN) is implemented to discriminate the signal process $tq\gamma$ from the background processes. This NN is trained on simulated data and receives characteristic event variables of $tq\gamma$ as input. Studying the significance and effects of different input parameters on the output may help narrow down or provide conditions for the event selection to optimise background suppression which would be useful for the differential analysis. Additionally, the investigation of these input features could provide vital insights into the nature of the $tq\gamma$ process. In this thesis, the correlations of characteristic features of $tq\gamma$ with the NN output are analysed. Subsequently, two input features, the transverse momentum of the photon p_T^γ and the sum of the forward jet energies and the photon energy, are further studied. These input features are divided into specific energy regions. The effects of different divisions on the NN output are then thoroughly examined and discussed.

2 Single top quark production with a photon in the Standard Model

2.1 A brief overview of the Standard Model

The Standard Model (SM) of particle physics, a quantum field theory, describes today's best theory of particle physics. In the SM, there are different kinds of elementary particles and three fundamental forces of nature: the electromagnetic force, the strong force and the weak force. Every force coincides with an elementary particle, called a boson, that acts as a mediator of the interaction. Another group of particles are called the fermions, and they only interact with these bosons if they have specific quantities, which are represented by their quantum numbers.

The fermions have spin $s = \frac{1}{2}$ and can be divided into two separate groups. The first group, named quarks, are colour charge carrying fermions. There are three up-type quarks (up, strange and top) with an electric charge of $q = +\frac{2}{3}e$ and three down-type quarks (down, charm and bottom) with an electric charge of $q = -\frac{1}{3}e$. The second group are the leptons. Three leptons (electron, muon and tau) have an electric charge of $q = +1e$. Furthermore, each of these leptons has a corresponding uncharged lepton partner called a neutrino. Three different families further categorize leptons and quarks. These quark and lepton families are ordered by mass and consist of an up-type quark, the corresponding down-type quark, a lepton and the corresponding neutrino. There is an anti-matter particle equivalent for all fermions where every charge-like quantum number has the opposite sign.

Particles with integer spin are called bosons. The first group of bosons, gauge bosons with spin $s = 1$, mediate the three fundamental forces. The gauge bosons with spin $s = 1$ are: *gluons* g, *photons* γ, Z and $W^{\pm}$. The Higgs boson H is the only boson with spin $s = 0$.

Gluons are colour charged and mediate the strong force between colour charged particles, including themselves. The six colour charges are *red*,*green*,*blue* and their anticolour counterpart. The strong force draws particles with colour together until a colour neutral state is achieved. This can occur when one quark bonds with a quark of the opposite colour, forming a meson. It can also occur when three quarks with *red*, *green*, *blue* colour charges respectively together form the colour neutral

Baryon. The strong force becomes stronger the further quarks are repelled from the colour neutral state. If quarks get repelled for a sufficiently high distance, two new quarks are formed, which then bond with the repelled. This process is called colour confinement and can occur many times in a row for higher energies, forming a shower of mesons and baryons. Photons mediate the electromagnetic (EM) between electrically charged particles. The massive bosons, Z and W^- as well as W^+ mediate the weak force. The weak force only acts on left-handed particles (and right-handed antiparticles). Here, left-handedness means that the spin direction is opposite to the direction of the momentum of the particle. Right-handed particles have their spin and momentum pointing at the same direction. The $W^\pm$ bosons are electrically charged, $q = \pm 1e$, and change the flavour of a quark when coupling to it. The flavour refers to the species of a fermion.

The Higgs boson arises from the electroweak theory. The Higgs mechanism provides an explanation for the presence of massive leptons and bosons by breaking the electroweak symmetry. The fermions acquire their mass by coupling with the Higgs boson via the so-called Yukawa interaction. Before the breaking of the electroweak symmetry, the gauge bosons exist as the electroweak eigenstates W^1, W^2, W^3 and B. The breaking of this symmetry mixes W^3 and B into the mass eigenstates Z and γ. The eigenstates W^1 and W^2 mix into the massive eigenstates W^+ and W^-.

An overview of the elementary particles in the Standard Model is given in figure 2.1.

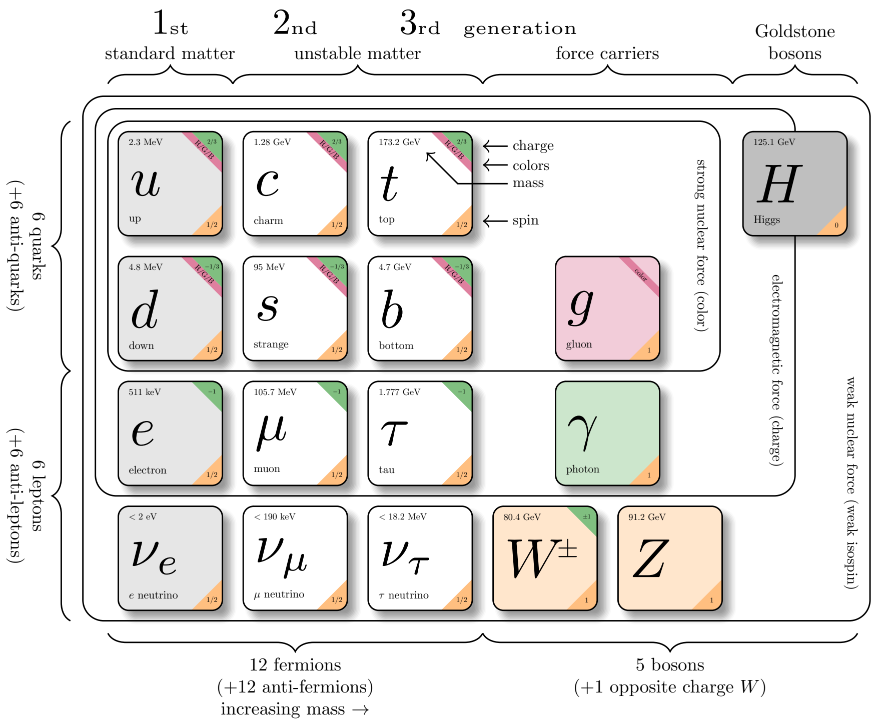


Figure 2.1: Elementary particles of the Standard Model alongside their properties [4].

2.2 The $tq\gamma$ process in the Standard Model

The top quark is an up-type quark and the most massive quark of the Standard Model with a mass of $m_t = 173.76 \pm 0.3\,\text{GeV}(S = 1.2)$ [15]. Additionally, the top quark has a very small decay width of $\Gamma = 1.42^{+0.19}_{-0.15}\,\text{GeV}(S = 1.4)$ [15] because of its high mass. This is one reason why the top quarks unlikely to build any bound states and always decay shortly after production. Only their decay products are observable and can be retraced back to the top quark.

Top quarks can be produced in three different channels: In the t-channel (tq), where a single top quark is produced when a bottom quark exchanges a W-boson with another quark, the s-channel (tb), where the top quarks are produced in top-antitop-pairs, and the tW-channel, where a gluon and a bottom couple and then decay into a single top and a W boson. In this thesis, the focus lies primarily on the t-channel production of the top quark. The top quark was first discovered in pair production at the Tevatron in 1995 during a proton-antiproton collision experiment (CITE). In 2009, the D0 [3] and CDF [2] collaborations also separately confirmed the observation of the t-channel top quark production at the Tevatron. The combined results are available in reference [9]. The CMS experiment at the Large Hadron Collider (LHC) of CERN [6] reported evidence for the t-channel single production of top quarks in association with a photon ($tq\gamma$) with a significance corresponding to $\sigma = 4.4$. The fiducial cross section was measured to be $\sigma(pp \to tq\gamma)(t \to \mu\nu b) = 115 \pm 17(stat) \pm 30(syst)\,\text{fb}$ for the photon transverse momentum $p_T^\gamma > 25\,\text{GeV}$ **??**.

For this thesis, the $tq\gamma$-events are produced in proton-proton-collisoins inside the ATLAS experiment. The ATLAS is discussed in detail in chapter 3. The production of processes in this experiment occurs with elementary particles inside of the protons, called partons. For the production of the $tq\gamma$-process, one gluon provided by the protons may produce a bottom-antibottom-quark pair. The bottom quark may then exchange a W-boson with a quark, turning the bottom quark into a top quark and changing the flavour of the quark. This top quark may then radiate a photon. It is essential to mention that while this thesis focuses on the top-photon vertex, the photon can be radiated from any charged particle elsewhere in the process. For instance, the bottom quark after the decay of the top may produce a photon. Afterwards, the top quark decays decays into a W^+-boson and a bottom quark. The W^+-boson then decays either into an antilepton and neutrino pair or a quark-antiquark pair of opposite quark types. However, only the leptonic decay mode is considered in this thesis.

In Figure 2.2 the leading order Feynman diagram for the $tq\gamma$ production is depicted. The charge conjugated diagram is not shown, but also considered in this thesis.

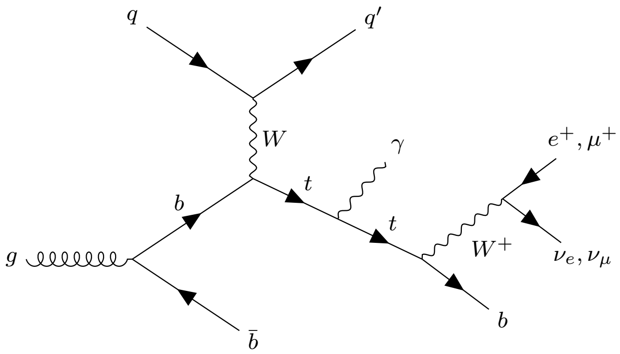


Figure 2.2: Leading order Feynman diagram of the $tq\gamma$ process in the Standard Model.

3 Measurement of $tq\gamma$

3.1 The ATLAS Experiment

The European Organization for Nuclear Research, known as CERN, located in Geneva, has various experiments studying elementary particles through the collision of heavy ions and protons. The Large Hadron Collider (LHC), the largest particle accelerator of CERN, has a circumference of 27 km and can collide particles with a center of mass energy of up to $\sqrt{s} = 13.6\,\text{TeV}$.

The LHC consists of four extensive experiments: the ALICE, the LHCb, the CMS and the ATLAS experiments. The research in this thesis is done with the help of the largest of these experiments, the ATLAS experiment. Figure 1 visualizes the structure of the ATLAS detector. A coordinate system needs to be defined in order to discuss the construction of the experiment. Three different coordinates are used to describe positions inside the experiment: First, the azimuthal angle ϕ, which ranges from 0 to 2π. Next, the pseudorapidity η, which is defined to be $\eta = -\ln(\tan\theta)$, where θ is the angle to the beam axis. The smaller θ is, the higher the pseudorapidity. And lastly, a distance ΔR, which can be defined in the ϕ-θ-plane as $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)}$.

The ATLAS detector is built symmetrically around the particle beam and can be divided into three subdetectors:

The inner detector tracks charged particles just after the collision. It consists of three different systems of sensors in a magnetic field parallel to the beam. These sensors are the pixel detector, the semiconductor tracker which works with silicone strips and a transition radiation tracker to track particles with gas-filled tubes.

In the EM calorimeter, metal layers (tungsten, copper or lead) absorb incoming particles and convert them into lower-energy particles called a shower. The calorimeters detect "showers" produced by electrons (and positrons), photons and hadrons. The barrel part of this calorimeter covers the pseudorapidity range $|\eta| <= 1.475$ and the end-cap components cover $1.375 < |\eta| < 3.2$. Hadrons do not deposit all of their energy into the EM calorimeter; their showers get absorbed by steel layers in the hadronic calorimeter behind the EM calorimeter. Here, plastic scintillating tiles produce photons that get converted into an electric current. The scintillating tiles cover the region $|\eta| < 1.7$. The region $1.5 < |\eta| < 4.9$ is then used by the cooper + liquid argon and tungsten + liquid argon calorimeter.

The muon spectrometer measures trajectories of muons with the help of a magnetic field. The spectrometer detects muons in the range of $|\eta| > 2.7$. Monitored drift tubes measure pseudorapidities up to $|\eta| = 2.0$ and cathode strip chambers cover higher pseudorapidities.

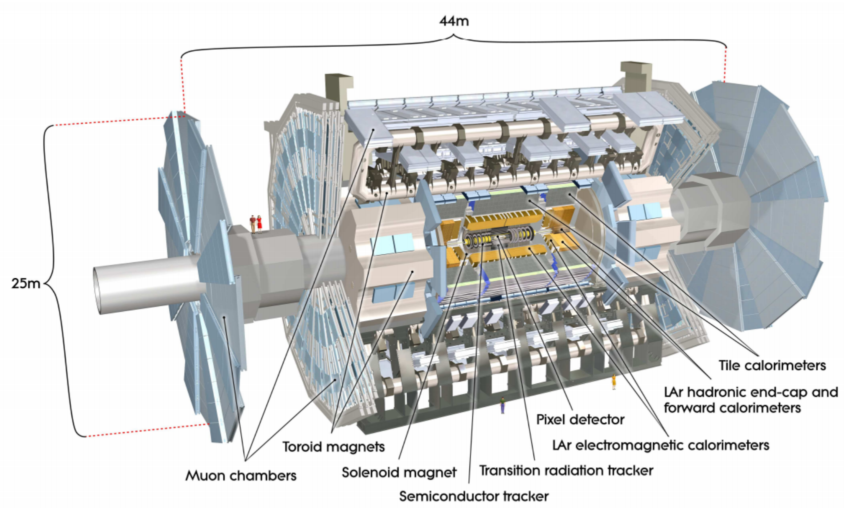


Figure 3.1: Schematic visualisation of the ATLAS Detector [8].

3.2 Object Reconstruction at the ATLAS experiment

3.2.1 Reconstruction of photons

Photons do not leave tracks inside the inner detector. They are reconstructed from clusters in the electromagnetic calorimeter. Photon candidates need to pass the *Tight* identification criteria with the transverse momentum being $p_T > 20\,\text{GeV}$ and $|\eta| < 2.3$. Photon candidtaes in the calorimeter transition region $1.37 < |\eta| < 1.52$ are exclused.

3.2.2 Reconstruction of leptons

Leptons like an electron also produce clusters inside the electromagnetic calorimeter and leave charged particle tracks in the inner detector. All leptons in the calorimeter region $1.37 < |\eta| < 1.52$ are excluded. Electron candidates have to pass the tight likelyhood identification (TightLH) conditions of $p_T > 27\,\text{GeV}$ and $|\eta| < 2.47$. To reconstruct muons, charged particle tracks in the inner detector are matched with muon spectrometer tracks. Muon candidates are required to pass the Medium identification with $p_T > 27\,\text{GeV}$ and $|\eta| < 2.5$.

3.2.3 Jets

Jets are showers of mostly mesons and fewer baryons in the hadronic calorimeter that result from the production of high energy quarks (colour confinement). They deposit some of their energy in calorimeter cells which then are combined into clusters by the anti-k_t algorithm [5] with a radius parameter of $R = 0.4$. It is then required that the cluster has a transverse momentum of $p_T > 25\,\text{GeV}$ and $|\eta| < 4.5$. If these conditions are met, the cluster is identified to be a jet.

Detector noise can lead to the misidentification of a jet. The nature of these misidentified jets has been studied thoroughly and a so-called ”jet cleaning procedure” is used to tag them. Any event containing at least one ”bad” jet is removed by the algorithm.

3.2.4 b-tagging

The identification of jets produced by bottom quarks (b-taggigng) must be made with high accuracy to distinguish *bottom* jets from *strange* jets. Here, the $DL1r$ algorithm is implemented for b-tagging. It is essentially a neural network trained on impact parameters and topological properties of decay vertices reconstructed within the jet. Only b-tagged jets passing 70% working point and $p_T > 20\,\text{GeV}$ are viewed as b-jets in this analysis. Detailed information on the $DL1r$-algorithm can be found in the references [10] and [11].

3.2.5 Missing transverse momentum E_T^{miss}

If all particle products are considered, there should be no magnitude for the sum of the transverse momentum p_T of all particles. Any measured magnitude is therefore attributed to an unmeasured particle. The missing transverse energy E_T^{miss} is consequently defined as the negative of this sum and assinged to a neutrino.

3.3 Background contributions from similar processes

Various processes besides $tq\gamma$ are also accepted by the criteria for event selection 4.2. However, for the scope of this thesis, contributions from these processes are considered background. The process $t\bar{t}\gamma$ holds the most similar decay product. Because of the second top quark, This process does have an additional b-jet, but it may not get b-tagged correctly. If the second weak decay also does not get correctly reconstructed, then the $t\bar{t}\gamma$ process virtually looks identical to $tq\gamma$. The $t\bar{t}\gamma$ process was found to have a cross-section of $\sigma(pp \to t\bar{t}\gamma) = 139 \pm 7\,(stat.) \pm 17\,(syst.)$ fb [1]. Next most similar processes are the production of a W-boson with jets, a Z-boson with jets and $t\bar{t}$.

Table 3.1 lists these and more of these processes contributing to the background.

	Process
1	$tq\gamma$
2	$t\bar{t}\gamma$
3	$W\gamma + jets$
4	$Z\gamma + jets$
5	$t\bar{t}$
6	s-$chan$
7	tW
8	t-$chan$
9	VV
10	$W + jets$
11	$Z + jets$

Table 3.1: List of SM processes that contribute to background noise in the measurement of $tq\gamma$.

4 Monte Carlo samples and event selection

4.1 Generation of Monte Carlo samples

The framework MadGraph5_aMC@NLO is used for Monte Carlo (MC) simulations of the considered $tq\gamma$ process. MadGraph5 is a matrix element generator that allows the interfacing of different packages for further simulation. The simulated events are generated at next-to-leading order (NLO) at the t-channel of single top production. The generator is interfaced to the package Pythia $v8.240$, which provides parton showers. The MadSpin and EvtGEN $v1.6.0$ packages give decay simulations of the top and bottom quark, respectively. Here, only leptonic decays of the top quark are considered.

Moving on to background processes, the $t\bar{t}$ process is modelled at leading order (LO) also using MadGraph5_aMC@NLO $v2.3.3$ interfaced to Pythia $v8.212$. Simulation of $W\gamma$+jets and $Z\gamma$+jets events are produced at NLO using the Shepra $v2.2.2$ and Shepra $v2.2.4$ packages. For the $t\bar{t}$ process and t-, s-, tW-channels Powheg-Box is used where Pythia $v8.230$ is again used as the showering program. The modeling here is performed in NLO in QCD.

The table 4.1 gives a summary of the generated samples and their generators.

Process	Generator
$tq\gamma$	*MadGraph5_aMC@NLO + Pythia8*
$t\bar{t}\gamma$	*MadGraph5 + Pythia8*
$W\gamma + jets$	*Sherpa* 2.2.2
$Z\gamma + jets$	*Sherpa* 2.2.4
$t\bar{t}$	*Powheg + Pythia8*
single top	*Powheg + Pythia8*
$W + jets$	*Sherpa* 2.2.1
$Z + jets$	*Sherpa* 2.2.1
Diboson	*Sherpa* 2.2.2

Table 4.1: List of generated samples alongside their generators.

4.2 Event selection

The selection criteria for events must hold the necessary conditions for a $tq\gamma$-process. It also needs to have enough restrictions to reduce background contributions as much as possible. Signal events have precisely one lepton, at least one photon and one b-tagged jet in the final state. The lepton should have a transverse momentum higher than 20 GeV, the photons momentum higher than 27 GeV and the b-tagged jet has to pass the $DL1r$-algorithm with a 70% working point.
Additionally, the missing transverse energy E_T^{miss} ought to be above 30 GeV to account for the neutrino in the decay mode. Finally, to reduce leading background contributions from the $Z \to ee(\to \gamma)$ process, the invariant mass of the leading photon and an electron candidate $m_{e\gamma}$ is set to be in the range $80\,\text{GeV} < m_{e\gamma} < 110\,\text{GeV}$. Altogether, this makes up the following requirements for selected events:

1. At least one photon γ with $p_T > 20\,\text{GeV}$
2. Exactly one lepton with $p_T > 27\,\text{GeV}$
3. $E_T^{miss} > 30\,\text{GeV}$
4. Exactly one b-tagged jet passing 70% working point (WP) of the $DL1r$-algorithm.
5. Invariant mass of leading photon and electron candidate between values $80\,\text{GeV} < m_{e\gamma} < 110\,\text{GeV}$

5 The Neural Network used for signal-background classification

5.1 Short introduction to neural networks

Neural networks are loosely on the human brain and are a means of doing machine learning, in which a computer learns to perform some task by analyzing training examples. A NN is usually organized into layers of processing nodes. These processing nodes are densely interconnected between layers, and every connection weighted. During the $forward\ propagation$, where the NN is tested on provided data, computations in the NN propagate from the input layer to the output layer. An individual node receives data from nodes in the layer beneath it and sends data to nodes in the layer above. Nodes multiply received data by their weight value and add them together to a single value. Only if the node exceeds a specific threshhold does it send its value to the next layer. During *backpropagation*, where the NN is being trained, weights and thresholds are continually adjusted until training data with the same label yield similar results. In this thesis, a NN discriminates between the $tq\gamma$ signal and background events. The NN is trained on Monte Carlo simulations (Section 4.1) and is tested on measurement data after that. Characteristic variables of events, which are discussed in detail in section 5.3, are used as the input.

5.2 The neural network architecture

The NN are built using the `Keras` library running on `TensorFlow` [7] [13]. Two different neural networks are trained to separate the signal from the background. One is trained on the zero-forward jet signal region (0-fj), and the other is trained on the At least one forward jet region ($\geq$ 1-fj). This is done to optimize the sensitivity of the analysis as the signal to background ratio (S/B) is greater for the $\geq$ 1-fj region than the 0-fj region.
Both models consist of one input layer, three densely connected node layers (Dense layer) and one output layer. The input layers have nodes for each input feature. The NN for 0-fj events has 16 input features while the NN for $\geq$ 1-fj events has

27 due to additional variables of the forward jets. The activation function for the Dense layers is the Leaky Rectified Linear Unit (ReLU) function $f(x)$:

$$f(x) = \begin{cases} x, & \text{for } x \geq 0 \\ 0.5x, & \text{for } x < 0. \end{cases}$$

For the output layer, the activation function is the sigmoid function $\sigma(x)$:

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$

Finally, the `Adam` algorithm is used as the optimizer for updating the weights in the NN [12]. Figure 5.1 displays the described architecture of the NN models.

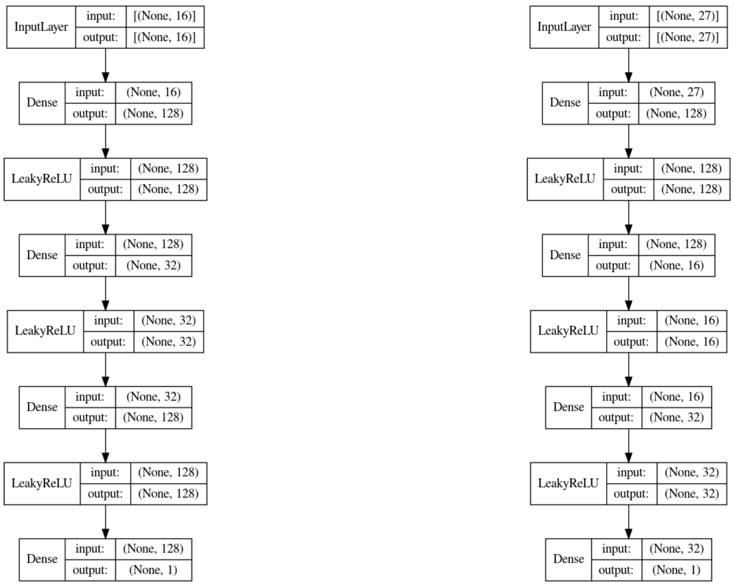


Figure 5.1: Visualization of the NN architecture for the zero forward jet region (left) and the ≥ 1 forward jets region (right).

5.3 Input features for the neural network

The input features used for the two NN models are listed in table 5.1 (Weiß nicht was ich hier schreiben sollte?).

	0fj variables	1fj variables
1	HT	HT
2	blep_dr	blep_dr
3	lbj_eta	lbj_eta
4	lbj_pt	lbj_pt
5	lbj_tagWeightBin_DL1r_Continuous	lbj_tagWeightBin_DL1r_Continuous
6	lep1_eta	lep1_eta
7	met_met	met_met
8	ph_pt	ph_pt
9	top_m	top_m
10	transMassWb	transMassWb
11	bph_pt	bph_m
12	topph_pt	topph_ctheta
13	ph_eta	ph_phi
14	lepph_dr	lep1_pt
15	transMassWph	met_phi
16	lep1_id	lbj_phi
17		Wbsn_e
18		bfj_m
19		blep_m
20		fj_eta
21		fj_phi
22		fjet_flag
23		fjph_ctheta
24		fjph_deta
25		fjph_dr
26		fjph_e
27		fjph_m

Table 5.1: **Variablen müssen in der mathematischen Darstellung angegeben werden!** Input variables of the NN trained on events with no forward jets and the NN trained on events with at least one forward jet.

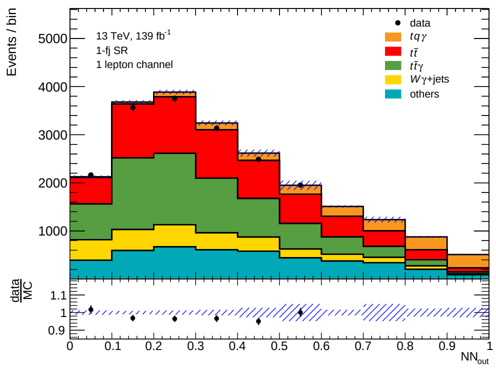
5.4 Performance and distribution of the NN output

A distribution for ten different bins of the NN output for $\geq 1\text{-}fj$ has been calculated and displayed in figure 5.2. In this plot, the contributions from different samples are labeled and stacked on top of each other. The data samples are viewed seperately and also added to the plot. To better visualize the composition of different bins, the right plot in figure 5.2 shows the percentage of each sample in each bin. These plots confirm that events at higher values of the NN output increase the S/B ratio. To determine the performance of the NN output, the signal efficiency is plotted alongside the background suppression (so-called "ROC-Curve") in figure 5.3. The signal efficiency (SE) is calculated as follows:

$$SE = \frac{\text{Amount of true signal events}}{\text{Total of signal events}} \text{Wie schreibe ich das besser?}$$

and the background suppression (BS) is calculated with the formula

$$BS = 1 - \frac{\text{Amount of true background events}}{\text{Total of background events}} \text{Wie schreibe ich das besser?}.$$



PLATZHALTER.
ABBILDUNG KOMMT NOCH

Figure 5.2: The NN output event distribution (left) and the composition of different bins of the NN output (right).

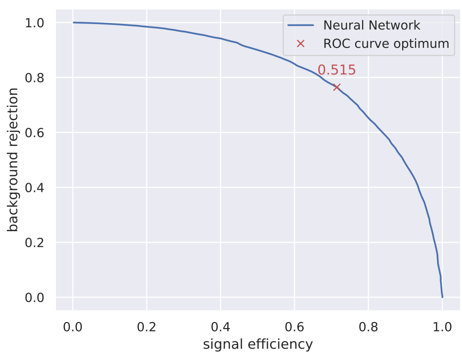


Figure 5.3: ROC-Curve of the neural network with the point of maximumized signal efficiency and backgrund rejection.

6 Differnetial analysis of the NN output

6.1 Correlations of input features with the NN output

6.2 NN output distribution dependence on photon p_T and fjet+photon energy

7 Conclusions

8 TO DO

1. Fix ugly citation style
2. Chapter: Neural Network
3. Chapter: Analysis
4. Chapter: Conclusion

Bibliography

[1] M. Aaboud et al. “Measurement of the t t ¯

$$t\bar{t}\gamma$$

production cross section in proton-proton collisions at s = 8

$$\sqrt{s} = 8$$

TeV with the ATLAS detector.” In: *Journal of High Energy Physics* 2017.11 (Nov. 2017). ISSN: 1029-8479. DOI: `10.1007/jhep11(2017)086`. URL: `http://dx.doi.org/10.1007/JHEP11(2017)086`.

[2] T. Aaltonen et al. “Observation of Electroweak Single Top-Quark Production.” In: *Phys. Rev. Lett.* 103 (9 Aug. 2009), p. 092002. DOI: `10.1103/PhysRevLett.103.092002`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.103.092002`.

[3] V. M. Abazov et al. “Observation of Single Top-Quark Production.” In: *Phys. Rev. Lett.* 103 (9 Aug. 2009), p. 092001. DOI: `10.1103/PhysRevLett.103.092001`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.103.092001`.

[4] Carsten Burgard. *Example: Standard model of physics*. Dec. 2016. URL: `https://texample.net/tikz/examples/model-physics/` (visited on 06/30/2021).

[5] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. “The anti-ktjet clustering algorithm.” In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: `10.1088/1126-6708/2008/04/063`. URL: `http://dx.doi.org/10.1088/1126-6708/2008/04/063`.

[6] S. Chatrchyan et al. “The CMS Experiment at the CERN LHC.” In: *JINST* 3 (2008), S08004. DOI: `10.1088/1748-0221/3/08/S08004`.

[7] Francois Chollet et al. *Keras*. 2015. URL: `https://github.com/fchollet/keras`.

[8] The ATLAS Collaboration et al. “The ATLAS Experiment at the CERN Large Hadron Collider.” In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: `10.1088/1748-0221/3/08/s08003`. URL: `https://doi.org/10.1088/1748-0221/3/08/s08003`.

[9] Tevatron Electroweak Working Group. *Combination of CDF and D0 Measurements of the Single Top Production Cross Section*. 2009. arXiv: 0908.2171 [hep-ex].

[10] *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-003. Geneva: CERN, Mar. 2017. URL: https://cds.cern.ch/record/2255226.

[11] *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-003. Geneva: CERN, Mar. 2017. URL: https://cds.cern.ch/record/2255226.

[12] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[13] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[14] A. M. Sirunyan et al. "Evidence for the Associated Production of a Single Top Quark and a Photon in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV." In: *Phys. Rev. Lett.* 121 (22 Nov. 2018), p. 221802. DOI: 10.1103/PhysRevLett.121.221802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.121.221802.

[15] P.A. Zyla et al. "Review of Particle Physics." In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.