TECHNISCHE UNIVERSITÄT WIEN

DIPLOMARBEIT

Recurrent neural networks for improving measurements with the top quark

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

im Rahmen des Studiums

Technische Physik

eingereicht von

Rosmarie Schöfbeck
Matrikelnummer 01525808

ausgeführt am Institut für Hochenergiephysik
der Österreichischen Akademie der Wissenschaften

Betreuung
Betreuer: Priv.-Doz. Dipl.-Ing. Dr.techn. **Robert Schöfbeck**

Ort, Datum	Betreuer	Verfasserin

Abstract

The Standard Model of Particle Physics (SM) is astonishingly successful in predicting fundamental particle physics. Experiments probing the Standard Model and extensions beyond the SM (BSM), are located at CERN's Large Hadron Collider (LHC), a particle accelerator situated at the French-Swiss boarder. The Compact Muon Solenoid (CMS), one of the four big experiments at the LHC, records proton-proton collisions at a center-of-mass energy of up to 13 TeV.

Data of the CMS experiment accumulated during the Run-II of the LHC, taking place from 2016 to 2018, is analysed in this thesis. This proton collision data is used to study different top quark processes, namely the $t\bar{t}Z$, the tWZ and the $t\bar{t}t\bar{t}$ signal processes. The latter two have not been observed yet and are promising targets for LHC Run III, yielding possible windows to BSM phenomena. There are many challenges in finding signal process signatures in the vast amount of data the CMS experiment provides. Often other, similar signatures give rise to huge backgrounds. Within this work, recurrent neural networks were used in order to improve signal-background discrimination in the named top quark processes.

Recurrent Neural Networks (RNNs), in particular Longtermshortmemory (LSTM) networks were used to separate the signal. Expected Limits were calculated and it was shown that LSTMs are efficient in slightly improving the limits of the different processes.

Kurzfassung

Das Standardmodel der Teilchenphysik (SM) ist erstaunlich erfolgreich in der Beschreibung fundamentaler Teilchenphysik. Am großen Hadronen Speicherring (Large Hadron Collider, LHC) am Europäischen Kernforschungszentrum CERN in Genf nehmen verschieden Experimente das Standardmodel genau unter die Lupe. Das CMS Experiment ist eines der vier größten Experimente und nimmt Proton-Proton Kollisionen bei Schwerpunktsenergien von bis zu 13 TeV auf.

Daten des CMS Experiments, welche während dem zweiten Lauf des LHC gesammelt wurden, werden in dieser Arbeit analysiert. Damit werden folgende Prozesse genauer untersucht: $t\bar{t}Z$, tWZ und $t\bar{t}t\bar{t}$. Letztere wurden noch nicht entdeckt, könnten aber mögliche Ziele für eine RunIII Wirkungsquerschnittsmessung sein. Die Messungen solcher Prozessen bergen einige Herausforderungen, beispielsweise werden gesuchte Signale oft mit ähnlichen Hintergrundprozessen überlagert. Im Zuge dieser Arbeit werden "Recurrent Neural Networks" (RNN) verwendet um die Trennung von Signal und Hintergründen zu verbessern.

Konventionelle neuronale Netzwerke sowie Longtermshortmemory (LSTM) Architekturen wurden für die Trainings zur besseren Isolierung des Signals verwendet. Limits wurden berechnet und es konnte gezeigt werden, dass LSTMs leicht verbesserte Resultate liefern, im Vergleich zu herkömmlichen neuronalen Netzen.

Acknowledgements

I want to thank all the people who supported me during my years at university. Working on this thesis in midst of a pandemic turned out to be extremely challenging, for various reasons. First and foremost I want to thank Robert for supervising this thesis and always finding a solution. Many thanks also to Lukas, who introduced me to the world of data analysis.

I want to express sincere gratitude to my parents, Sophie and Erich, for supporting me at all times. Also, special thanks to my sisters Anna and Flora, my grandparents and all of my family, whom I can always count on. I am grateful for the good friends who life would be boring without; the adventures, the ups and the downs we experience together.

My biggest thank you is for Daniel, who listens to, comforts and challenges me as well as, most importantly, has a superb sens of humor.

Contents

1 The Standard Model

The standard model of particle physics (SM) [1–4] is a theory describing all known elementary particles as well as their interactions. It is a mathematical theory that predicts the behaviour of fundamental particles up to the GUT scale and is tested up to the TeV scale. Despite its astonishing success in providing experimental predictions, the SM cannot explain various phenomena that could be a sign of physics beyond the Standard Model (BSM). In the following, an introduction to the SM is given and its shortcomings are discussed briefly.

1.1 Fundamental Interactions and Particles

The SM includes three out of four fundamental interactions, namely the weak-, strong- and electromagnetic force. The impact of the not included gravitational interaction is very weak on particle scale. In Table 1.1, an overview on the fundamental interactions and their characteristics can be found.

Fundamental particles are point-like objects which carry momentum and spin. They can furthermore have different charges associated with the fundamental interactions. The electric charge corresponds to the electromagnetic interaction, the weak charge to the weak interaction and the color charge to the strong interaction. Figure 1.1 illustrates the main interactions of the SM as well as its particle content.

The fundamental particles contained in the SM can be classified according to their spin into fermions and bosons.

Bosons are integer spin particles and are described by Bose-Einstein statistics. Vector bosons

Table 1.1: The fundamental Interactions of the Standard Model.

Interaction	**Boson**		**Spin**	**Mass [GeV]**	**Strength**
Strong	Gluon	g	1	0	1
Electromagnetic	Photon	γ	1	0	10^{-3}
Weak	W		1	80.4	10^{-8}
	Z		1	91.2	

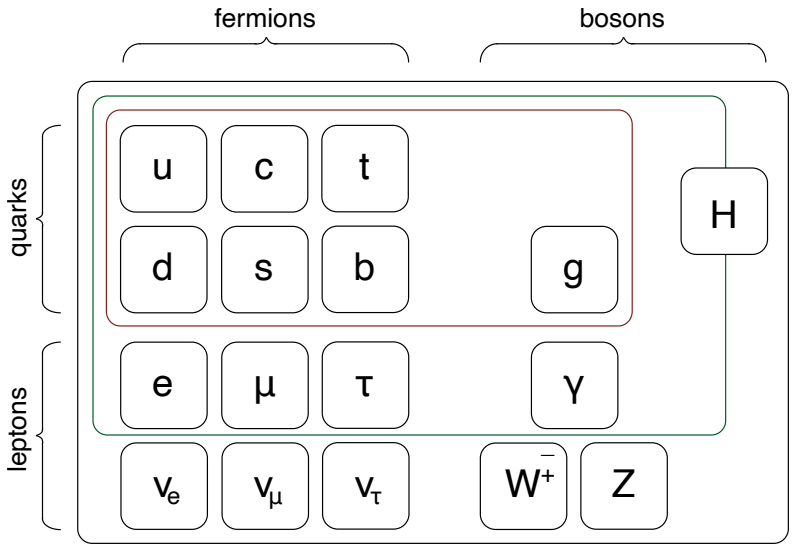

Figure 1.1: Overview of SM particle content [5]. Lines indicate groups of particles that interact in the SM to leading order via the strong and the electroweak force and the Higgs boson.

Table 1.2: Fermion properties.

	Leptons			**Quarks**		
	Name	Charge	Mass/GeV	Name	Charge	Mass/GeV
1. Gen	electron	-1	0,0005	up	$-1/3$	0,003
	electron-neutrino	0	$\leq 2\times10^{-9}$	down	$2/3$	0,005
2. Gen	muon	-1	0,105	strange	$-1/3$	0,1
	muon-neutrino	0	$\leq 2\times10^{-9}$	charme	$2/3$	1,3
3. Gen	tau	-1	1,78	bottom	$-1/3$	4,5
	tau-neutrino	0	$\leq 2\times10^{-9}$	top	$2/3$	174

have spin 1 and mediate the SM forces. As shown in Table 1.1 the photon is the force carrier of the elctromagnetic interaction, gluons mediate the strong force and the massive Z and W bosons mediate the weak force. Gluons can self-interact since they carry color charge.

The only spin-0 boson, the Higgs Boson (H), was jointly discovered in 2012 by the CMS [6] and the ATLAS [7] experiments. The mass of fundamental particles is proportional to their coupling to the H boson.

Fermions are half-integer spin particles grouped in quarks an leptons. They obey the Fermi-Dirac statistics. There are 12 fermions in three generations, which are listed in Table 1.2. All fermions have corresponding anti-particles.

Charged leptons interact via electromagnetic and weak interaction. The neutrinos are neutral leptons that only interact via the weak force. The first generation charged lepton, the electron, is a building block of the atom.

Quarks carry color charge, fractional electric charge and weak charge and therfore participate in all SM interactions. Protons and neutrons are three-quark bound states of quarks and gluons (hadrons), which are the constituents of the atomic nuclei. Particles made up of one quark and one anti quark are called mesons.

The mathematical formulation of the Standard Model is based on quantum field theory. A detailed description can be found eg. in [1, 2, 8].

The SM is described by the local gauge symmetry group

$$\mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y \tag{1.1}$$

where each interaction is associated with a gauge symmetry of the total symmetry group: the strong force correspons to $\mathrm{SU}(3)_\mathrm{C}$ and the electroweak force to $\mathrm{SU}(2)_\mathrm{L} \otimes \mathrm{U}(1)_\mathrm{Y}$. The electroweak interaction breaks into the electromagentic and weak interaction at lower energies.

The SM is described mathematically by its Lagrangian density, a function of the fields and their derivatives. Field excitations correspond in this context to the previously described particles of the SM. The Euler-Lagrange equations for the scalar fields ϕ_i can be written as

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)} \right) - \frac{\partial \mathcal{L}}{\partial \phi_i} = 0 \tag{1.2}$$

where $\mathcal{L}$ is the Lagrangian density. The Lagrangian L of a system is the spacial volume integral over $\mathcal{L}$

$$L = \int \mathcal{L}(\phi, \partial_\mu \phi) d^3x. \tag{1.3}$$

The electroweak interaction

The electromagnetic interaction is mediated by the photon. It acts upon particles carrying electric charge. In the following a short summary of the mathematical description is given in order to obtain the Lagrangian.

Using the Euler Lagrange equations 1.2 and the Lagrangian density for free fermions

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{1.4}$$

the Dirac equation can be derived:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0. \tag{1.5}$$

It describes the behaviour and properties of fermions using Dirac Spinors ψ and $\overline{\psi} = \psi^\dagger \gamma^0$. The Lagrangian density is required to be invariant under local phase transformations U(1)

$$\psi(x) \rightarrow \psi'(x) = \psi(x) e^{iq\xi(x)}. \tag{1.6}$$

which is ensured by using the covariant derivative D_μ instead of ∂_μ

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iqA_\mu. \tag{1.7}$$

The newly introduced field A_μ is the photon field. In QED the Lagrangian density contains the electromagnetic field strength tensor $F_{\mu\nu}$ and describes the fields of the electron, the photon and their interactions and can be written as

$$\mathcal{L}_{QED} = \overline{\psi}(i\gamma^\mu D_\mu - m)\psi + e\overline{\psi}\gamma^\mu \psi A_\mu - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{1.8}$$

For the mathematical formulation of the weak interaction the Lagrangian is required to be invariant under $SU(2)_L$ transformations. Hereby, the derivative is again replaced with the covariant derivative D_μ

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig_W \mathbf{T} \cdot \mathbf{W}_\mu(x) \tag{1.9}$$

which contains the three generators of the group $\mathbf{T} = \frac{1}{2}\sigma$ as well as the three new gauge fields $\mathbf{W}(x)$ corresponding to the gauge bosons of the weak interaction W^+,W^- and Z^0. The weak interaction is the only parity conservation violating interaction and uniquely able to change quark flavours. All SM particles interact weakly.

The weak and electromagnetic interactions can be described in a unified theory introduced by Salam, Weinberg and Glashow [3][4][9] during the 1960s. Hence, the electromagnetic and weak interaction are different aspects of the same interaction. The symmetry group of the electroweak interaction is $SU(2)_L \otimes U(1)_Y$ and its covariant derivative is given by

$$D_\mu = \partial_\mu + \frac{ig_W}{2}\mathbf{T}\cdot\mathbf{W}_\mu + \frac{ig'}{2}B_\mu \tag{1.10}$$

where B_μ corresponds to U(1) and $\mathbf{W}$ to SU(2) as mentioned above. The physical gauge bosons are products of spontaneous symmetry breaking, which is described shortly later on, and are linear combinations of the fields:

$$W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2) \tag{1.11}$$

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W^3 \end{pmatrix} \tag{1.12}$$

where θ_W is the weak mixing or Weinberg angle.

The strong interaction

The only participants in the strong interaction are quarks. As suggested by its name, it is the strongest interaction and mediated by gluons. Gluons carry color charge enabling them to self interact. In the theory of the strong interaction, also called quantum chromodynamics (QCD) there a three color states: red, green and blue. QCD's symmetry group is $SU(3)_C$ and its generators are the Gell-Mann matrices T^a. In accordance with the eight generator matrices one needs to introduce eight gauge fields G^a_μ (a = 1, ..., 8)

$$\mathbf{G}_\mu = \sum_{a=1}^{8} G^a_\mu T_a. \tag{1.13}$$

The covariant derivative can be defined as

$$D_\mu = \partial_\mu + ig_S\mathbf{G}_\mu \tag{1.14}$$

containing the strong coupling constant g_S.

Quarks cannot be observed freely due to confinement, a characteristic of QCD. Only colorless particles, consistent of three particles carrying different color charge (hadrons) or a pair of a particle and its antiparticle (mesons), can be observed.

Spontaneous symmetry breaking

Brout, Englert and Higgs developed a theory to describe the acquisition of mass of the elementary particles[10][11]. The so called Englert-Brout-Higgs mechanism (EBH) incorporates symmetry breaking in the SM and proposes the existence of the Higgs boson. For every broken symmetry group generator there exists a so called Goldstone boson, a massless field. From Goldstone's theorem [12] spantaneous symmetry breaking in particle physics was developed. It states that die Lagrangian densitiy is invariant under symmetry transformations, whereas the vacuum value (lowest energy value) is not, which is described in the following.

After introducing a complex scalar field ϕ

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.15}$$

one can write down its Lagranian density:

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi). \tag{1.16}$$

Here, D_μ is again the covariant derivative and V is the Higgs potential of the form

$$V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{1.17}$$

The minimum of the potential corresponds to the vacuum expectation value of the filed. As sketched in Figure. 1.2 the minimum for $\mu^2 > 0$ is trivial, whereas there are infinite minima for $\mu^2 < 0$, indicated by the dashed circle.

The minima are defined by

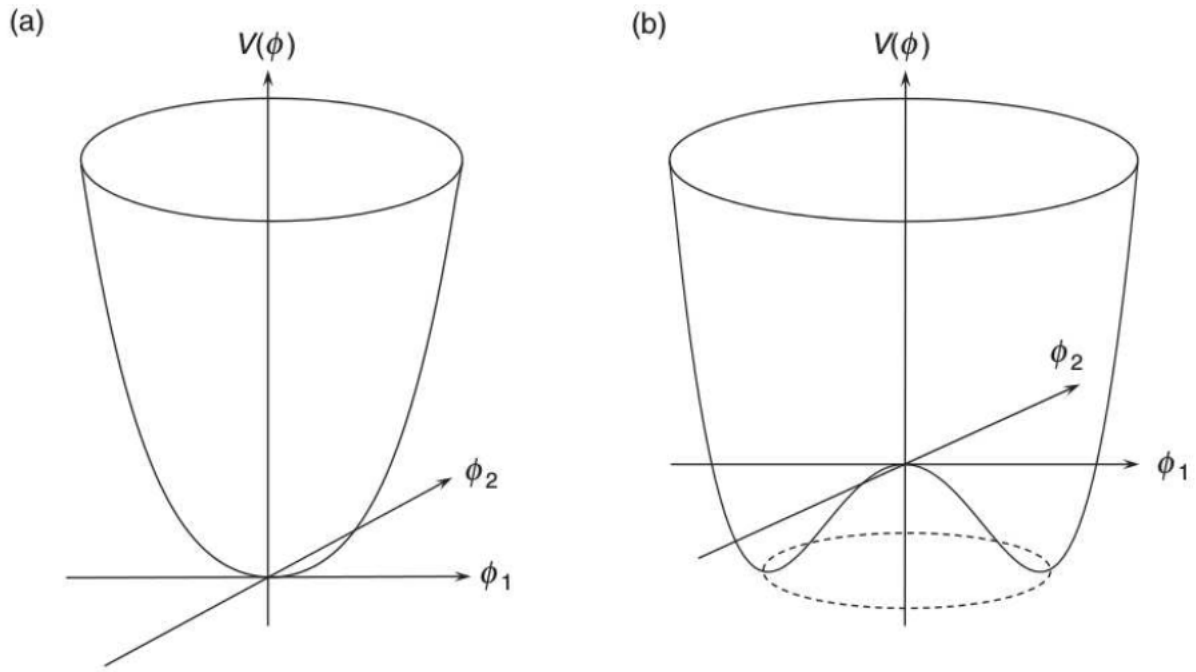

Figure 1.2: The potential V(ϕ) for a)$\mu^2 > 0$ and b) $\mu^2 < 0$ [1].

$$\phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda} = v^2 \tag{1.18}$$

and the physical vacuum state v can be *chosen* which corresponds to breaking the symmetry group. The field is expanded around v and the Goldstone bosons can be gauged away using the *unitary gauge*. The chosen, now entirely real, scalar field can be written as

$$\phi(x) = \frac{1}{\sqrt{2}}(v + h(x)) \tag{1.19}$$

where h is the physical Higgs field.

Using the unitary gauge, the fields in the Lagrangian correspond to physical particles, however, the physical predictions are independed of the chosen gauge. Inserting the Field in Eq. 1.19 into the Lagrangian density in Eq. 1.16 the following Lagrangian can be gathered.

$$\mathcal{L} = \overbrace{\frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \frac{1}{2}m_H^2 h^2}^{\text{massive h scalar}} + \overbrace{\frac{1}{2}m_W^2 W^{\mu+}W_\mu^- + \frac{1}{2}m_Z^2 Z^\mu Z_\mu}^{\text{massive gauge bosons}} + \overbrace{V_{int}}^{\text{interaction terms}} \tag{1.20}$$

It describes the massive Higgs field h and gauge bosons as well as interaction terms V_{int} including boson interactions and self couplings of h. Therefore spontaneous symmetry breaking in the SM is induced by the Higgs vacuum expectiation value (VEV) and responsible for the fermion and non-vanishing boson masses [13]. Explicit terms for the boson masses can be expressed as followed:

$$m_H^2 = 2\lambda v^2 \tag{1.21}$$

$$m_W^2 = \frac{g_W^2 v^2}{4} \tag{1.22}$$

$$m_Z^2 = (g_W^2 + g'^2)\frac{v^2}{4}. \tag{1.23}$$

.

Similarly, one can obtain a general fermion mass term [14]

$$m_f = \frac{g_f v}{\sqrt{2}}. \tag{1.24}$$

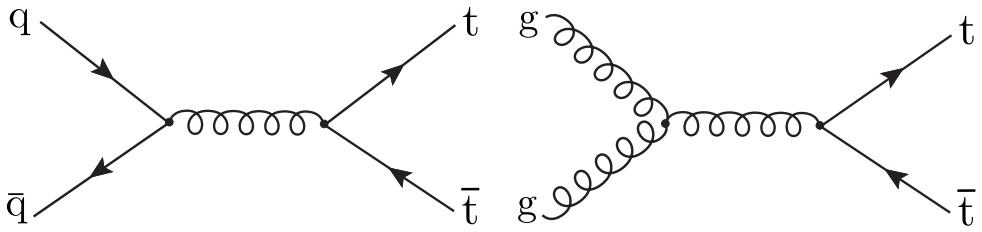

Figure 1.3: Exemplary leading order Feynman diagrams of top quark pair production [18]. Left: Production by $q\bar{q}$ annihilation. Right: Production by gluon-gluon fusion.

1.2 Top Quark Physics

The top (t) quark was discovered at the Tevatron accelerator at Fermilab in 1995 [15] using antiproton-proton collision data. It is the heaviest particle with a mass of 172.44 ± 0.49 GeV [16]. The lifetime of the t quark is around 5×10^{-25} s which is smaller than the QCD interactions timescale [17]. As a consequence the t quark decays before hadronization.

Top quarks are mainly produced in top-antitop $t\bar{t}$ quark pairs at the Large Hadron Collider (LHC). In the priority of the cases (85%) the production happens via quark-antiquark annihilation $q\bar{q} \rightarrow t\bar{t}$, whereas 15% are produced by gluon-gluon fusion $gg \rightarrow t\bar{t}$. These production types via the strong interaction are indicated in Figure 1.3.

It is also possible to produce a singe t quark via the weak interaction. Examples of Feynman diagrams for t- and s-channel production are shown in Figure 1.4.

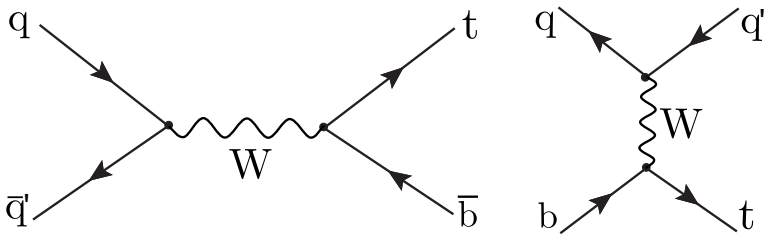

Figure 1.4: Exemplary leading order Feynman diagrams of single top quark production. Left: Production in the s-channel. Right: Production in the t-channel.

In the SM a t quark decays into a W boson and a b quark in 99,8% of all cases, because CKM factors suppress other decay modes.

Since the t quark is the heaviest particle, it has the largest coupling to the H boson and is therefore predestined to test the Higgs sector. Precise measurement of its properties also

Table 1.3: Vector boson decay modes [21].

	decay modes	**Fraction** Γ_i/Γ
W^+	$l^+\nu$	$(10.86 \pm 0.09)\%$
	$e^+\nu$	$(10.71 \pm 0.16)\%$
	$\mu^+\nu$	$(10.63 \pm 0.15)\%$
	$\tau^+\nu$	$(11.38 \pm 0.21)\%$
	hadrons	$(76.41 \pm 0.27)\%$
Z	l^+l^-	$(3.3658 \pm 0.0023)\%$
	e^+e^-	$(3.3632 \pm 0.0042)\%$
	$\mu^+\mu^-$	$(3.3662 \pm 0.0066)\%$
	$\tau^+\tau^-$	$(3.3696 \pm 0.0083)\%$
	invisible	$(20.000 \pm 0.055)\%$
	hadrons	$(69.911 \pm 0.056)\%$

probes symmetry breaking and electroweak interactions [19]. The anomalous interactions of the t are a window to new physics, a deviation from theoretical predictions can indicate BSM phenomena.

1.3 Vector boson decays

In particle physics the branching ratio describes the number of particles decaying via a specific decay mode with respect to the total number of particles which decay via all decay modes [20]. The t quark decays with 99.8% nearly exclusively into a W boson and a b quark.

W bosons decay in approximatly 2/3 of all cases to hadrons and 1/3 to leptons. Z boson decays result again mostly in hadrons (70%), leptons (10%) or invisible decays (20%). Those are neutrinos, not seen by the the detector and therefore "invisible". More detailed numbers of decay modes are listed in Table 1.3 [21].

1.4 Limitations of the Standard Model

The Standard Model of Particle Physics has some deficiencies, despite being a most successfully theory. Examples of those shortcomings are described briefly in the following.

Neutrino Masses

It is known that neutrinos have mass, contrary to the SM prediction. Neutrinos can oscillate into their different flavours [22], which requires them to have mass. A possible solution for the generation of neutrino masses is the see-saw mechanism [23].

Parameters

Furthermore, 27 parameters have to be inserted to the theory because they cannot be predicted within the SM. These include the lepton and quark masses as well as the Higgs mass, the Higgs vacuum expectation value, the QCD vacuum angle and gauge couplings [24].

Dark Matter

Moreover, the SM is valid for baryonic matter, which makes up only 5% of our universe. The other 95% constitute of dark matter (DM) and dark energy [25]. DM does not interact electromagnetically and hence, no direct observation could be made so far. However, gravity does affect DM and indirect observations motivate its the existence: gravitational lensing effects, the cosmic microwave background as well as rotational properties of galaxies.

Gravity

Finally, the SM does not contain gravity. Physicists aim to combine General Relativity and the Standard Model in one theory, however, so far attempts have not been successful. However, since gravitational interaction is significantly weaker than the other fundamental interactions it can be neglected in particle physics experiments at CMS.

2 The Large Hadron Collider

The Large Hadron Collider (LHC) [26] is a particle accelerator and storage ring located near the French-Swiss boarder at the European Organization for Nuclear Research(CERN). The LHC has a circumference of 27 km and is used to collide beams of protons or heavy ions. During Run II, taking place from 2016-2018, a center of mass energy for proton proton collisions of 13 TeV was reached; characteristic Figures of the LHC Run II can be found in table 2.1.

As depicted in Figure 2.1, pre-accleration is necessary in order to reach those high energies at the LHC. Firstly, protons produced from hydrogen ionization are accelerated in Linac2, a linear accelerator. After reaching an energy of 50 MeV, the protons are injected into synchotrons, the first beeing the Proton Synchotron Booster (PBS). At the next stage, protons will be accelerated to 25 GeV in the Proton Synchotron. The Super Proton Synchotron (SPS) is the predecessor of the LHC and acclerates protons to 450 GeV. It is interesting to note at this point, that at the SPS experiments UA1 and UA2 the W and Z bosons were discovered [27][28][29]. Finally, each beam is accelerated to 6.5 TeV at the LHC.

Superconducting magnets are used to focus and lead the particle beams on their circular paths. Before collision the two beams travel in opposite directions in separated beam pipes. There are seven experiments aiming to collect data and analyse the produced particles. There are four interaction points, each hosting one of four main experiments at the LHC:

- A Torroidal LHC ApperatuS (ATLAS) [31]
- A Large Ion Collider Experiment (ALICE) [32]
- Compact Muon Solenoid (CMS) [33]

Table 2.1: LHC Run II characteristics.

Quantity	**Number**
Circumference	26659 m
Dipole operating temperature	1.9 K
Number of magnets	9593
Number of main dipoles	1232
Number of main quadrupoles	392
Number of RF cavities	8 per beam
Nominal energy, protons	6.5 TeV
Nominal energy, ions	2.56 TeV/u
Nominal energy, protons collisions	13 TeV
Number of bunches per proton beam	2808
Number of protons per bunch(start)	1.2×10^{11}
Number of turns per second	11245
Number of collisions per second	1 billion

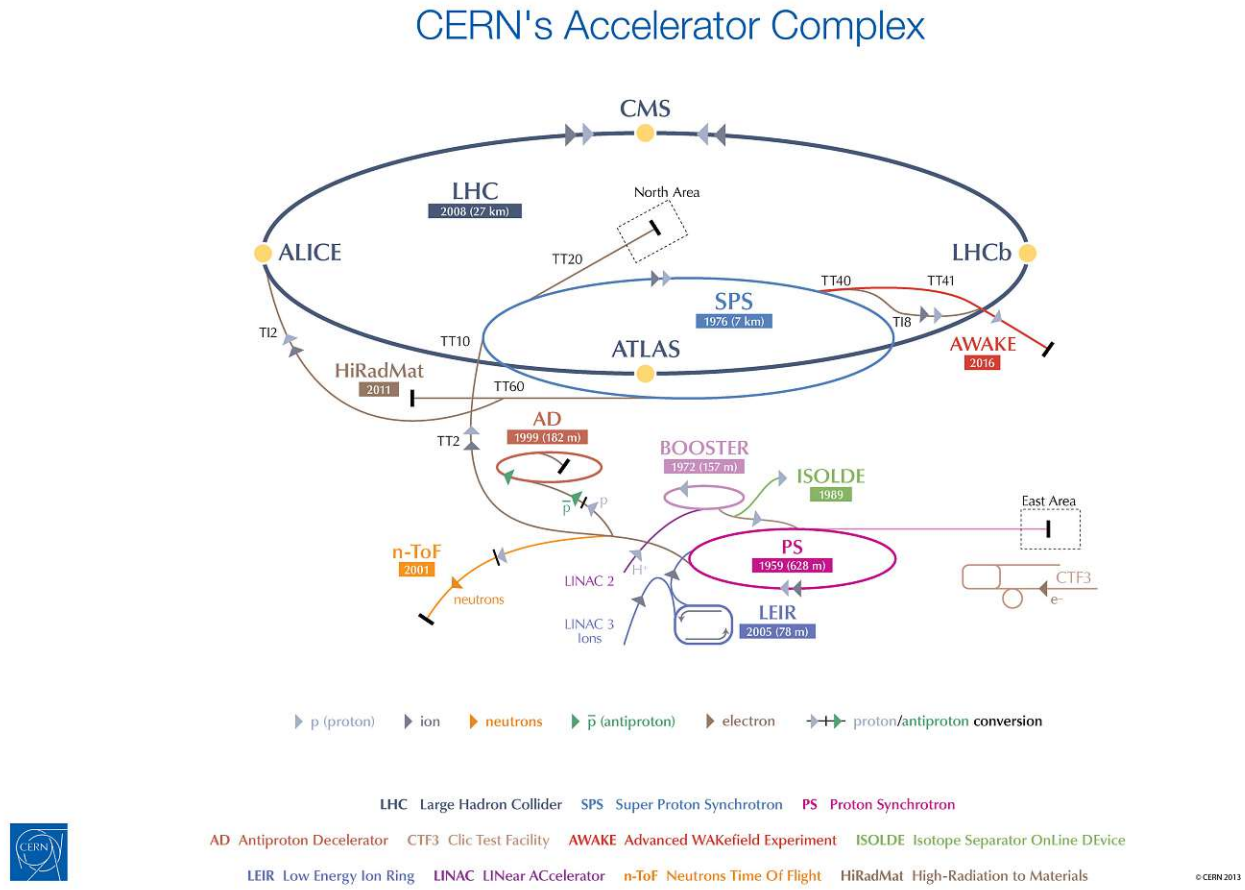

Figure 2.1: A sketch of CERN's accelerator complex. The four main experiments as well as pre-accelerators are shown, from [30].

- LHC beauty (LHCb) [34]

ATLAS and CMS are general purpose detectors. They aim to observe a broad range of physics and discover new phenomena. A more detailed overview of the CMS experiment will be given in the next chapter. The detectors of ATLAS and CMS differ in their designs, with ATLAS being the largest particle detector ever constructed.

ALICE is designed to investigate lead ion collisions and study quark-gluon plasma. In producing ion collisions, the LHC tries to mimic conditions in the very early universe. Extremely high temperatures lead to the production of quark-gluon plasma.

The LHCb detector is specialised in B-mesons, which are mesons containing an anti b quark, and made to find out more about antimatter-matter asymmetries. It detects mainly forward particles in contrast to ATLAS and CMS.

The LHC is among the most powerful particles colliders worldwide. Luminosity is, apart from beam energy, used to quantify the performance of a collider. It determines the number of interactions and links the rate of events and the cross section [35]

$$\frac{dR}{dt} = \mathcal{L}\sigma_p. \tag{2.1}$$

It can be calculated using

$$\mathcal{L} = \frac{N_b N_1 N_2 f}{4\epsilon\beta^*} \tag{2.2}$$

where N_1 and N_2 are the number of particles in each bunch, N_b the number of bunches, f the frequency, ϵ the normalized beam emittance and β^* the beta function at the collision point. At the LHC instantaneous luminosities of $\mathcal{L} = 10^{34}\text{cm}^{-2}s^{-1}$ are achieved. However, this also gives ground to the so called pileup effect. Several collisions take place per bunch crossing, which complicates the reconstruction and procession of collision products.

3 The CMS Experiment

The Compact Muon Solenoid (CMS) [36, 37] is a particle detector at the LHC. It is set up around 100 m underground and designed to observe proton-proton collisions. A broad range of measurements can be carried out with data collected by the CMS. A 3D sketch of the general-purpose detector is shown in Figure 3.1. It is 15 m high, 21 m long, weights 14000 tons and contains several layers of specially designed components to determine particle properties. Figure 3.2 depicts a slice trough the detector showing all its main components: the innermost silicon tracker, the electromagnetic and hadronic calorimeters, a superconducting solenoid magnet as well as a system to detect muons.

3.1 Coordinate System

A sketch of the coordinate system can be found in Figure 3.3. Its center lies in the collision point, the z-axis points in beam direction and the x-axis points inside the accelerator ring. The angle ϕ lies in the x-y plane whereas θ is the angle to the z-axis. It is convenient to use the pseudo rapidity

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{3.1}$$

instead of θ since $\Delta\eta$ is Lorentz invariant. Hence, there is a (along the z-direction) boost-

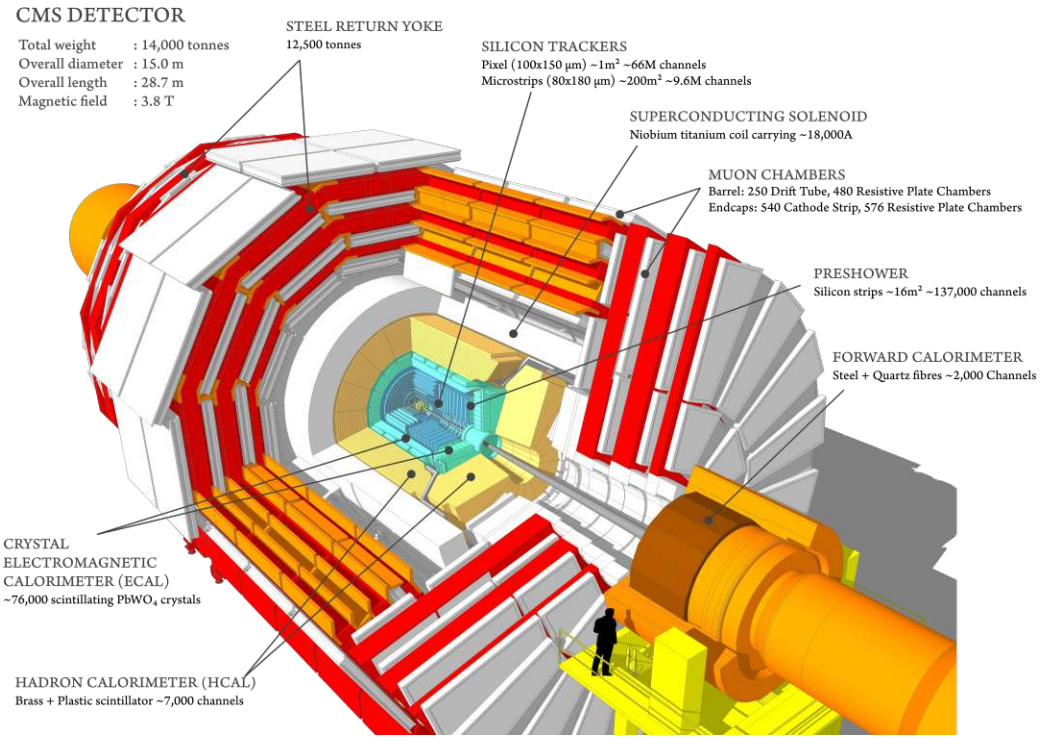

Figure 3.1: A sketch of the CMS detector layout [37].

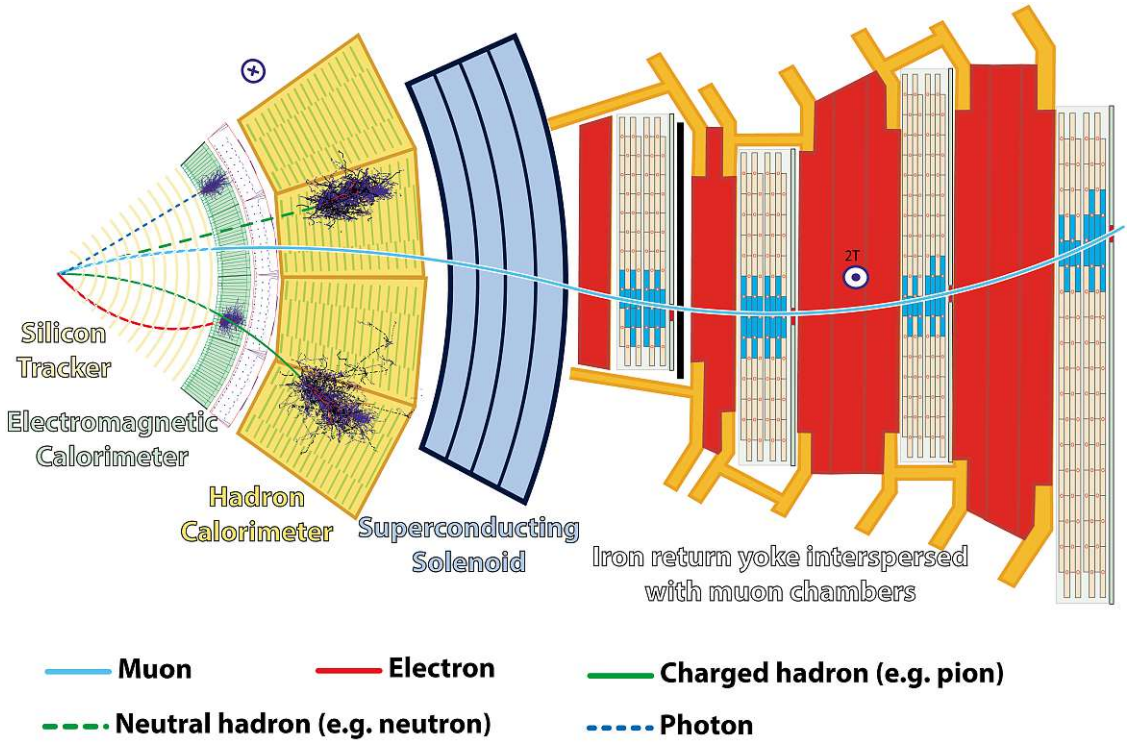

Figure 3.2: A slice trough the CMS detector showing the sub-detectors and how particles may interact with them [38].

invariant angular distance defined by

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{3.2}$$

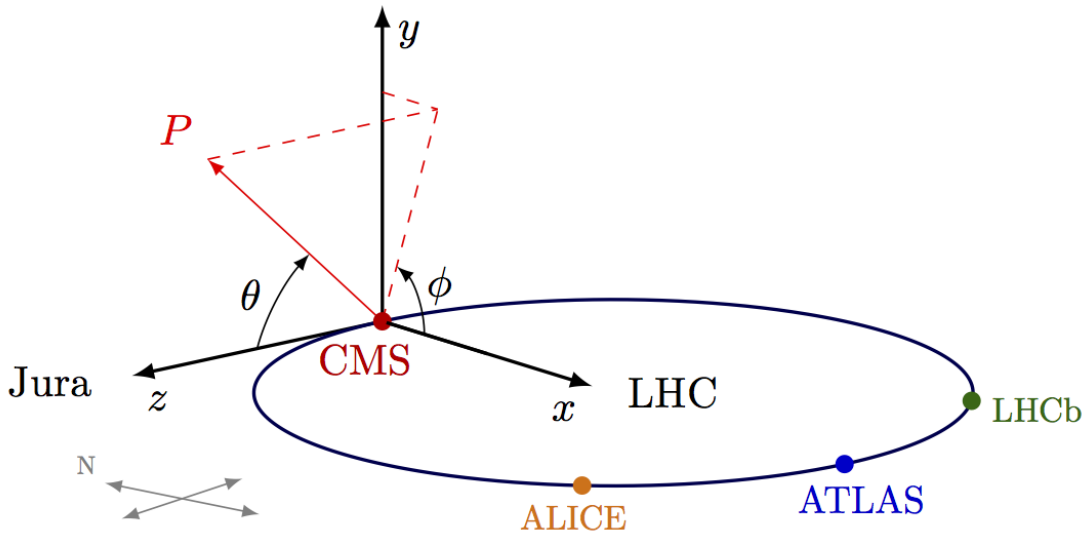

Figure 3.3: The Coordinate System [39].

3.2 Superconducting Solenoid Magnet

A naming part of CMS is the most powerful solenoid magnet, generating a 3.8 Tesla magnetic field [40]. It is made of superconducting NbTi, cooled to 4.45K and weights 12000t. The magnet is used to bend the paths of charged particles. Knowing their trajectories leads to identification of their charge as well as the measurement of the momentum of the charged particles.

3.3 Tracker

The paths taken by the particles are identified by a silicon Tracker made of millions of individual electronic sensors. It is the innermost layer of the detector and built around the collision point. The Tracker length is 5.8m and its diameter is 2.5m. It consists of silicon

pixel and -strip sensors which are highly resistant to radiation damage. They are arranged in concentric layers and silicon interacts electromagnetically with the passing particles. Each measurement is accurate to 10µm and they can then be joined together to reconstruct a particle track. The reconstruction of paths of high-energy muons, electrons and hadrons is possible.

3.4 Calorimeters

The particle energies are measured in two Calorimeters. The Electromagnetic Calorimeter (ECAL) measures electron and photon energies by stopping them, whereas Hadrons pass the ECAL and are only stopped in by the Hadron Calorimeter (HCAL) located further from the collision point.

3.4.1 Electromagnetic Calorimeter

The ECAL is divided in a barrel and two endcap sections. The former contains 61 200 lead tungstate crystals whereas the latter each contain additional 7324 crystals each. Furthermore, preshower detectors are located before the endcaps in order to distinguish between single high-energy photons and neutral pions. The underlying concept of the ECAL is electromagnetic showering: photons/electrons produce a shower of lower energy photons,electrons and positrons by pair production/bremstrahlung. The resulting particles can ionize atoms in the scintillator crystals if their energy is below a certain threshold. As a result the crystals will produce detactable light.

3.4.2 Hadron Calorimeter

The HCAL is a sampling calorimeter made of alternating layers of dens brass absorber plates and plastic scintillators. An interaction of a hadronic particle with the absorber material produces secondary particles. Those pass scintillators and additional absorber layers resulting in hadronic showers detected with light coming from the scintillators. This light is lead

to readout boxes using wavelength shifting fibres and is then amplified by photodetectors. The HCAL is structured into barrel, endcap and forward regions in order to detect the large showers. All but the outer barrel (HO) are located inside the magnet coil, which ensures the detection of all energies.

3.5 Muon System

Finally, muons can also be observed by a dedicated detector system. Since they are around 200 times heavier than electrons they are not stopped by the calorimeters. Their identification and measurement of momentum takes place in specially built sub-detectors, located at the edge of the experiment. The in total 1400 muon chambers include drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). DTs are installed concentrically around the beam line while CSCs are used in both endcaps. RPCs are very fast and provide a good temporal resolution and are therefore used in both, barrel and endcaps. They can differentiate between bunch crossings and are also utilised as triggers.

3.6 Trigger

Each beam contains 2808 bunches of about 10^{11} protons. The LHC bunch spacing is 25ns, so that the total $3{\times}10^{14}$ protons in the ring collide with a rate of 40MHz. This generates a tremendous amount of data, and a trigger system needs to be introduced [41]. Its goal is to select potentially interesting events and therefor reduce the amount of data. To achieve this, a good timing resolution and synchronisation is necessary.

A two level trigger system is set up for the preselection. The Level 1 trigger (L1) reduces the initial data by a factor of 1000 using only information provided by the calorimeters and the muon system. It reconstructs the signals to physical objects with field-programmable gate arrays. Those reconstructed objects (particle candidates) then get ranked and forwarded via global triggers to the High Level Trigger (HLT). The HLT uses full event data and attenuates the output rate to about 100 Hz. This amount of detector date can then be stored and used for analysis.

4 Machine Learning

Machine Learning (ML) has revolutionized interpreting large amounts of accumulated data and helps to increase discovery potential of experiments [42]. It is used in various areas of high energy physics and the number of different methods and available tools increases steadily [43]. Particle physics and computing advance hand in hand, building upon one another's successes.

ML is a sub-field of Artificial Intelligence (AI), which aims to mimic human cognitive functions. It is designed to use data to improve its algorithms, which in the end should assign the correct label to an unknown input [44]. For the learning process a subset of the available data (training data) is used. The quality of the ultimate model depends on its performance on previously unseen data (test data). Within this project, classifying ML techniques were used to distinguish signal from backgrounds.

4.1 Artificial Neural Networks

An Artificial neural networks (ANN) is a subset of ML. It is comprised of interconnected nodes (or neurons), each generating a response to a given input. The assembly of nodes mapps input variables to an output option [45]. There can be more than one layer of neurons, Neural Networks are therefor multi-layered structures of algorithms.

Each node is comprised of a set of inputs x_n, corresponding weights w_n, a bias (a constant value) and an output. A neuron is activated if a certain threshold is surpassed by the neuron activation potential. The activation potential p is the sum of the weighted inputs plus the

bias b:

$$p = \sum_n w_n x_n + b. \tag{4.1}$$

Applying an activation function to the potential gives to output o: $o = f(a)$.

A schematic sketch of one neuron is depicted in Figure 4.1.

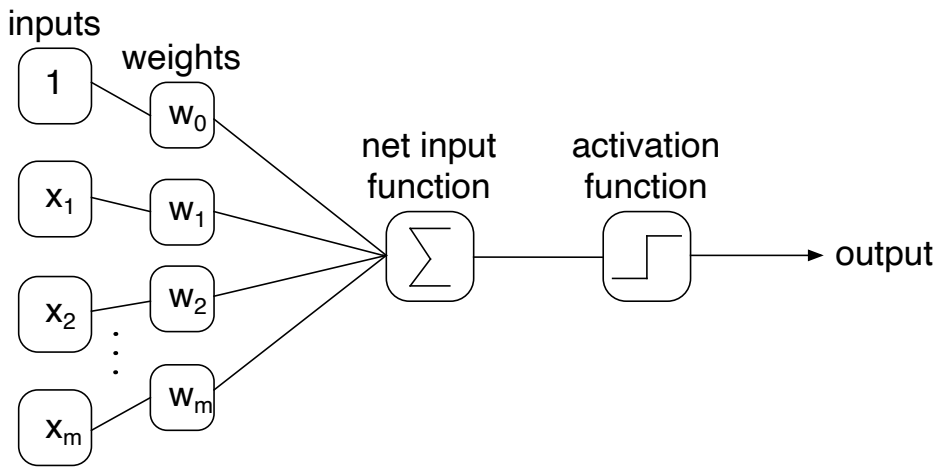

Figure 4.1: Cartoon of the possible architecture of one node.

Many functions can serve as activation functions, commonly the signum function, Heaviside function or hyperbolic tangent are used [46].

Not only one, but several layers of neurons add up to a Neural Network(NN). First there is an input layer, it contains the features of the used data which shall be taken into account. The last layer is the output: there is one output neuron for every class we consider. In between there can be several so-called "hidden layers". One speaks of deep neural networks [31] if there is more than one hidden layer, as depicted in Figure 4.2. In order to train a network, error signals need to be produced and back propagated. This results in two basic signal flows: forward propagation of function signals and back propagation of error signals [47], as illustrated in Figure 4.3.

For the error function, often a simple square error is chosen:

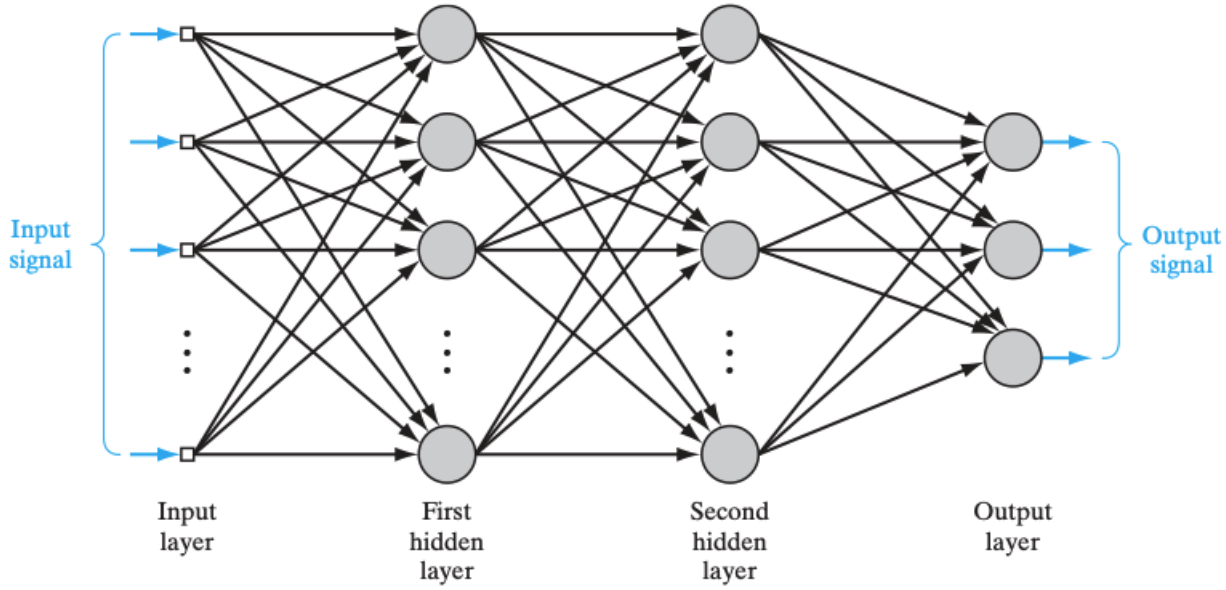

Figure 4.2: Graph of a multilayer perceptron with two hidden layers from [47].

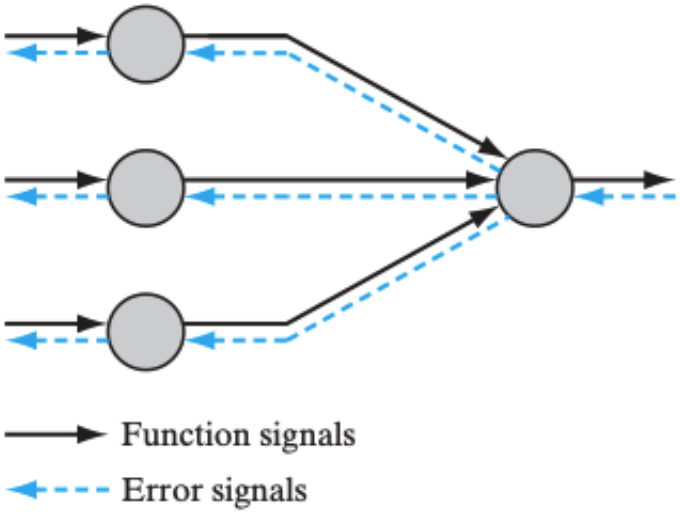

Figure 4.3: Illustration of directions of basic signal flows in a NN, from [47].

$$E(\vec{w}) = \frac{1}{2}\sum_{n}(t_n - o(x_n))^2. \tag{4.2}$$

To minimize the error, gradient descent is often the first choice. In this approach, each weight is adjusted as follows

$$\vec{w}(t+1) = \vec{w} - \gamma \nabla E(\vec{w}(t)). \tag{4.3}$$

Here, γ is the learning rate and t is the training step. The first needs to be chosen individually for each problem. All examples in a training sample make up an epoch. When batch learning is used, weights are adjusted after each epoch. For networks of neurons, which might have several layers, the back-propagation algorithm is needed. It is necessary to calculate the derivatives of the error function 4.2 with respect to the weights. This is realized by applying the chain rule

$$\frac{\partial E_i}{\partial w_{nj}} = \frac{\partial E_i}{\partial p_n}\frac{\partial p_n}{\partial w_{nj}}. \tag{4.4}$$

Opposed to a single neuron, neural networks are able to tackle non-linear problems. In this thesis RNNs are used for multi-class classification tasks.

4.1.1 Recurrent Neural Networks

Recurrent neural networks (RNNs) posses at least one feedback loop. Neurons get information from previous time steps. RNNs are very powerful because they can store information for a short period of time and it is possible to process inputs of variable length. They are used to learn long term dependencies [48]. A basic RNN sturcture containing a hidden layer and loops is shown in Figure 4.4.

However, due to their gradient-based learning technique the gradients are prone to explode or vanish after several time steps [48], the so called vanishing gradient problem [49]. In addition, it is difficult to store information [50] for a longer period of time. This leads to the

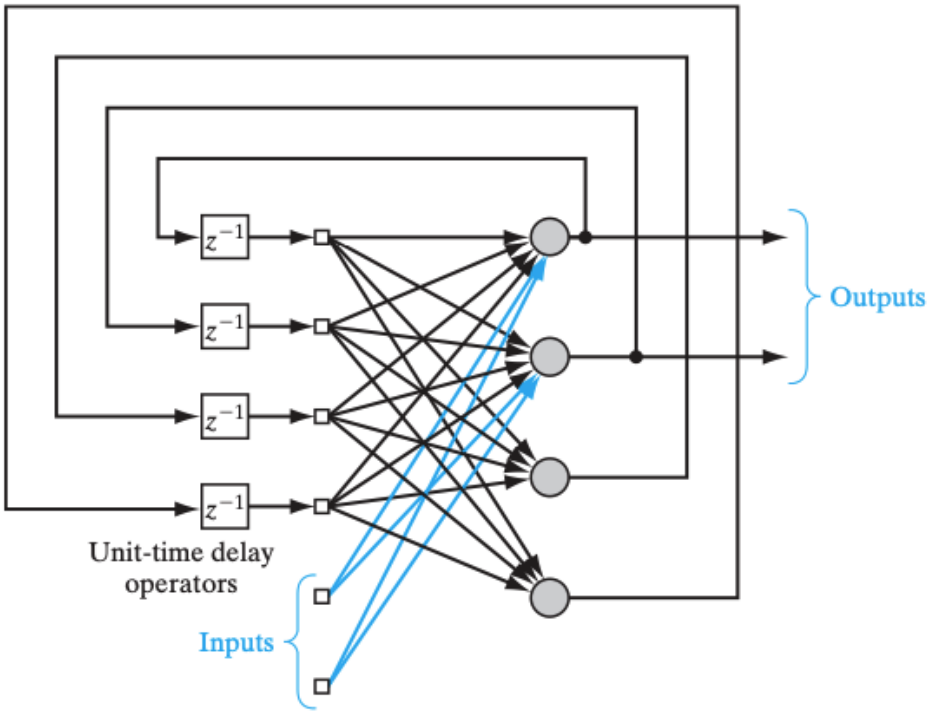

Figure 4.4: Graph of a basic RNN structure containing loops [47].

invention of long short-term memories.

Long Short Term Memory

Long short-term memories (LSTMs) were already introduced in 1997 [51]. They are an approach to solve the problem of information storage in RNNs. The main idea is the usage of memory cells. It is important to note that LSTMs avoid the vanishing gradient problem which occurred in previously discussed RNNs.

The memory cell contains three gates: input, output and forget gate. Figure 4.5 shows a LSTM memory cell where z,i,f,c, o and y are vectors describing the following

z: cell input activations

i: input gate activations

f: forget gate activations

c: memory cell states

i: output gate activations

f: cell output activations

G is the activation function for the cell input, h for the cell state and σ for the gates. The

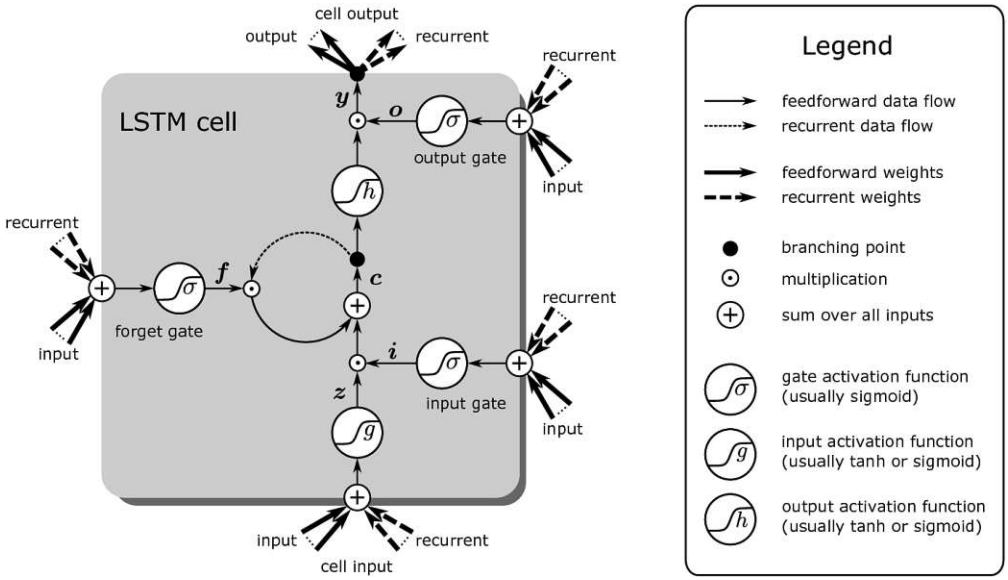

Figure 4.5: A LSTM memory cell described in the main text [52].

structure of memory cells allows LSTMS for *uniform credit assignment*, meaning errors can be back-propagated to neurons without scaling, opposed to standard RNNs. Inputs get assigned the same errors regardless of the time the input was first considered. This is a huge advantage of LSTMs since every input feature can be considered at each training step.

The usage of LSTM networks has vastly increased since the 1990s. The application fields are extremely diverse and many works have been published on this topic [53]. The structure has proven especially successful in language applications [52] like speech transcriptions or machine translations. Companies like Google, Apple and Microsoft use LSTM for their speech recognition programs. Networks are trained with huge amounts of data, using the advancements in graphic card technologies to perform several matrix multiplications simultaneously.
Very interestingly, LSTM applications include environmental aspects like prediction in water

qualtiy [54] and medicine [55], even translating neural activity into speech [56]. The overwhelming success of the LSTM network architecture has not yet reached the world of particle physics. Hence why the appliability to complex classification problems is investigated in this thesis.

In the scope of this work KERAS software framework [57] as well as the Toolkit for Multivariate Data Analysis with ROOT(TMVA) [45] were used for training.

5 Analysis

To identify different particles and processes, one searches the data for decay signatures. In this thesis Monte-Carlo (MC) samples are used for all backround and signal samples. Those were produced privately. In the following each of the investigated processes are described and signal region plots are included. In addition, parameters for the used machine learning algorithms are reported.

5.1 The tWZ process

In the tWZ process, we consider the production of a t quark as well as a W and a Z boson. The process has not been observed so far, but is a promising target for the LHC Run-III. Representative Feynman diagrams are depicted in Figure 5.1. The signal has a t quark, a Z boson and a W boson in its final state, as shown in the left and middle diagrams. On the right side, there is one extra jet. This can be used to study the top Yukawa coupling also present in the right diagram [58]. We consider the 3l channel of the tWZ process. Hence, we require three leptons. The Z boson can decay leptonically into two leptons as described above, and the W boson can decay into one lepton and one neutrino. To ensure one Z boson, we again require an opposite sign same flavour lepton pair which is within a window of ± 15 GeV of the Z boson mass. By requiring at least three jets, including one b jet, ttZ and WZ become the most dominant backgrounds.

The main backgrounds to the tWZ process are the $t\bar{t}Z$ and the WZ processes. The latter is described below, whereas the former is described in section 5.3. The $t\bar{t}Z$ final state is very similar to the tWZ final state, differing only by one additional b quark.

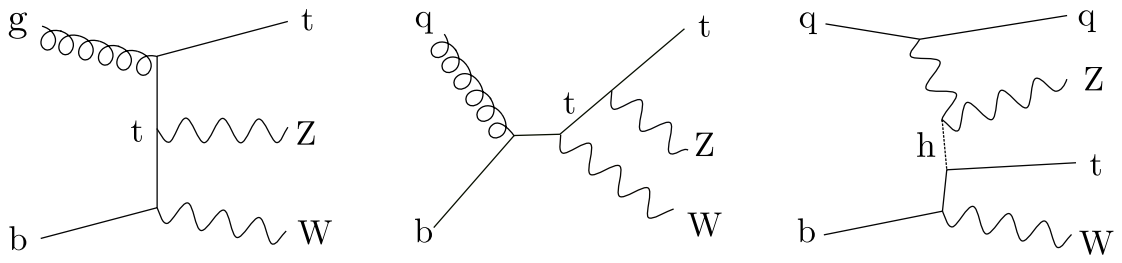

Figure 5.1: Representative tWZ tree level Feynman diagrams.

The WZ production in pp collisions has been measured by CMS and ATLAS experiments [29]. The leptonic decay channel is therefor well understood. The massive gauge boson pair decays to three leptons and one neutrino: WZ→ $l\nu l'l'$ with $l, l = e, \mu$.

5.2 The $t\bar{t}t\bar{t}$ process

Similar to the tWZ process, the $t\bar{t}t\bar{t}$ process has not been discovered yet, though predicted by the SM [59]. Figure 5.2 shows a 4t Feynman diagram. The signal contains to $t\bar{t}$ quark pairs.

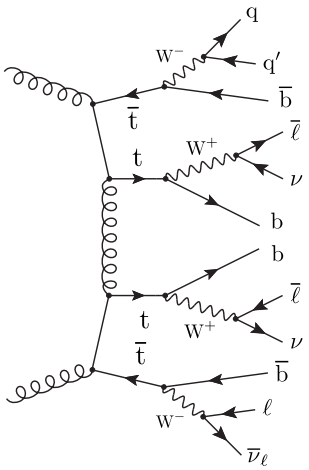

Figure 5.2: Representative $t\bar{t}t\bar{t}$ tree level Feynman diagram for production in the $3l$ decay channel.

In this thesis, we are looking at a three lepton decay of the $t\bar{t}t\bar{t}$ process. The lepton pair is off Z, meaning that its energy is not close to the Z mass. In its pure 3l final state, there are four bjets and four W bosons. Those decay either leptonically or hadronically, resulting in

3l and 2 jets. The detector is, again, not perfect. Hence, we require at least 4 jets, of which two are be b tagged.

The main backgrounds to the $t\bar{t}t\bar{t}$ process are $t\bar{t}W$, $t\bar{t}Z$ as well as nonprompt leptons. The $t\bar{t}Z$ process is described in 5.3 in more detail. In the SM, the production of a top quark pair and a W boson is rare [60]. It yields different decay signatures due to it's many final state particles. Hence, it is a main background in the search for two top quark pairs. Processes with large cross sections, such as Drell Yan, mainly give rise to nonprompt backgrounds. Nonprompt leptons do not originate in the four top decay, but rather from jets or are so called fake leptons. Fake means the detector miss identified jets as leptons.

5.3 The $t\bar{t}Z$ process

The $t\bar{t}Z$ process holds an antitop top quark pair and a Z boson in its final state. The CMS experiment already measured the $t\bar{t}Z$ cross-section at a center of mass energy of 13 TeV. This resulted in a cross-section of $\sigma(t\bar{t}Z) = 0.95 \pm 0.05\ (stat) \pm 0.06\ (syst)$pb [61].

Figure 5.3 shows examples for leading order feynman diagrams which yield the $t\bar{t}Z$ process. The top left diagram depicts an initial state radiation, where the Z boson is irradiated before the interaction. The other processes are final state radiation processes, characterised by a Z boson radiated by scattered particles, in our case top quarks.

In this thesis we are looking at a two lepton decay of the $t\bar{t}Z$ process. The Z boson decays into two leptons of opposite sign (OS) and the top quarks dacay hadronically. Considering top quark and Z boson branching ratios described previously 1.21.3 one can compose the signal composition as shwon in Figure 5.4. The two OS lepton channel constitutes 10.2 % of the total $t\bar{t}Z$ events. It is required that the mass of the dilepton pair is close to the Z boson mass, to select Z bosons. The two leptons are charged, have the same flavour and the opposite sign. The top quark will produce three jets including one b tagged. This results in a final state of six jets, including two b quarks and two leptons. Since, again, there might be mistakes as a result of the detector performance, a looser selection is required for the analysis. We require therefor at least 5 jets, with at least one b jet.

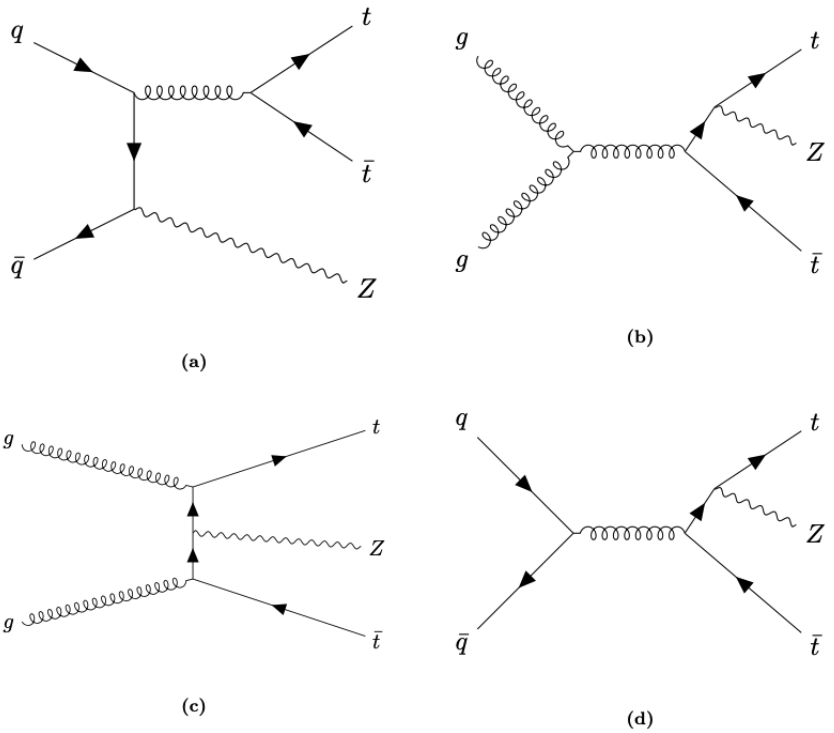

Figure 5.3: Representative $t\bar{t}Z$ tree level Feynman diagrams.

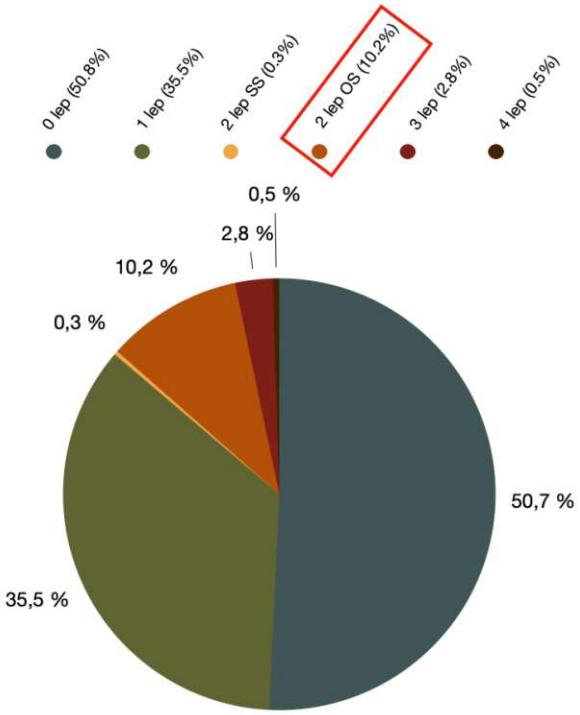

Figure 5.4: The distribution of different $t\bar{t}Z$ final states in a pie chart. [62]

There are processes with a similar detector signature to the $t\bar{t}Z$ process. To distinguish the signal from those backgrounds, machine learning techniques will be used. The main backgrounds to $t\bar{t}Z$ are Drell-Yan (DY) and $t\bar{t}$ production.

A lot of top-antitop quark ($t\bar{t}$) pairs are produced, but to be of concern to $t\bar{t}Z$ events both W bosons need to decay leptonically. Similar to the $t\bar{t}Z$ signal signature, the W bosons decay in lepton pairs of opposite sign. The requirement of the dilepton mass beeing close to the Z mass supresses W decays into same flavour leptons, which could be mistaken as Z bosons. In addition, the W decay produces neutrinos, whereas in the $t\bar{t}Z$ process no real MET is expected. Figure 5.5 shows a possible Feynman diagram for $t\bar{t}$ quark production.

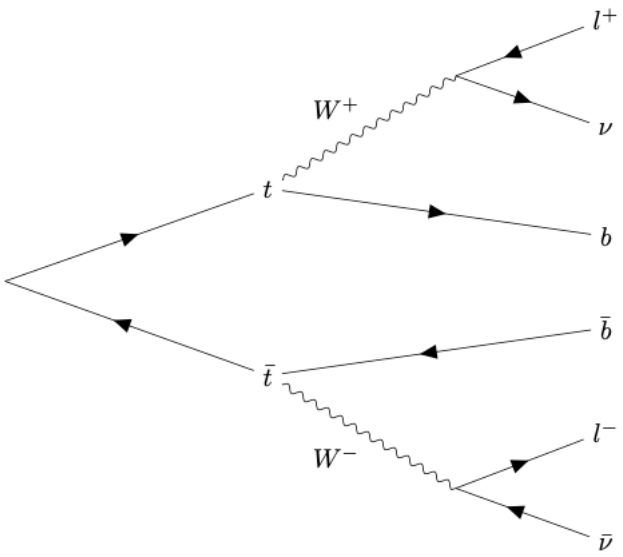

Figure 5.5: A Feynman diagram for the production of $t\bar{t}$ background.

In the Drell-Yan (DY) process a Z boson and a virtual photon are produced through qark-antiquark annihilation. The DY Feynman diagram is depicted in Figure 5.6. Although there are no obvious jets in this process, gluons radiated in the initial state can form jets. Unfortunately DY has a very large cross section. Together with the Z boson, the signature can then look similar to $t\bar{t}Z$.

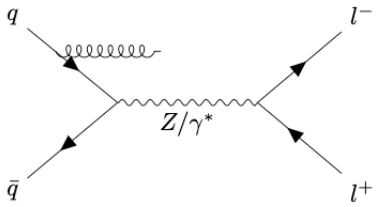

Figure 5.6: A Feynman diagram for Drell-Yan background.

5.4 Training variables

The goal of this thesis was to use ML techniques in order to improve the separation of signals and backgrounds. For the training configuration different variables for each process are used. The variables were specifically selected. For the tWZ process an extensive study was conducted to find the combination of variables giving the best results. For each process, all used properties are listed in Appendix table 1. The Machine Learning Algorithms are provided with different lepton and jet kinematics as well as global event properties. The individual variables are described in the following. Figures 5.7-5.9 show signal region plot for different variables for each process. The ratio depicts the ratio between signal and backgrounds.

Transverse momentum p_T is the momentum component perpendicular to the beamline. Since the overall momentum is not known in particle collisions one sticks to the transverse plane, where the momentum before collision knowingly equals zero.

Pseudorapidity as described in section 3.1, the pseudorapidity η is commonly used in high energy physics to account the angel of a particle relative to the beam axis:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{5.1}$$

$\Delta\mathbf{R}$ is the boost invariant angular distance defined as:

$$\Delta\mathrm{R} = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{5.2}$$

Azimuth angle Phi is the angle from the x-axis as shown in Figure 3.3.

A Jet is composed of different particles originating in hadronization processes. Good jets are such ones that are not fake or mismeasured. If more than one jet is required they are numbered according to the highest p_T. B-tagged jets are also an important variable, because the number of bs and other jets differs in each process.

Missing Energy Neutrinos only interact weakly and leave no sign in the detector. Hence, they cannot be measured directly. By calculating the transverse momentum of the event, neutrinos can be measured indirectly corresponding to energy imbalance in the transverse plane:

$$\vec{p}_T^{\,miss} = - \sum_{particles} \vec{p}_T. \tag{5.3}$$

Not in all processes missing energy is expected, therefor it can be used as discriminating variable.

H_T refers to the sum of the p_T of all jets:

$$H_T = \sum_{jets} p_T. \tag{5.4}$$

Transverse mass The transverse mass m_T for two objects a and b is defined as follows

$$m_T = \sqrt{2 p_T(a) p_T(b)[1 - \cos\Delta\phi(a,b)]}. \tag{5.5}$$

For a Z boson cadidate the two lepton p_Ts are used while a W candidate requires one lepton p_T and missing momentum. The transverse mass is also boost invariant. Of course, the concept can be extended to calculate the 3 lepton invariant mass (m_{3l}).

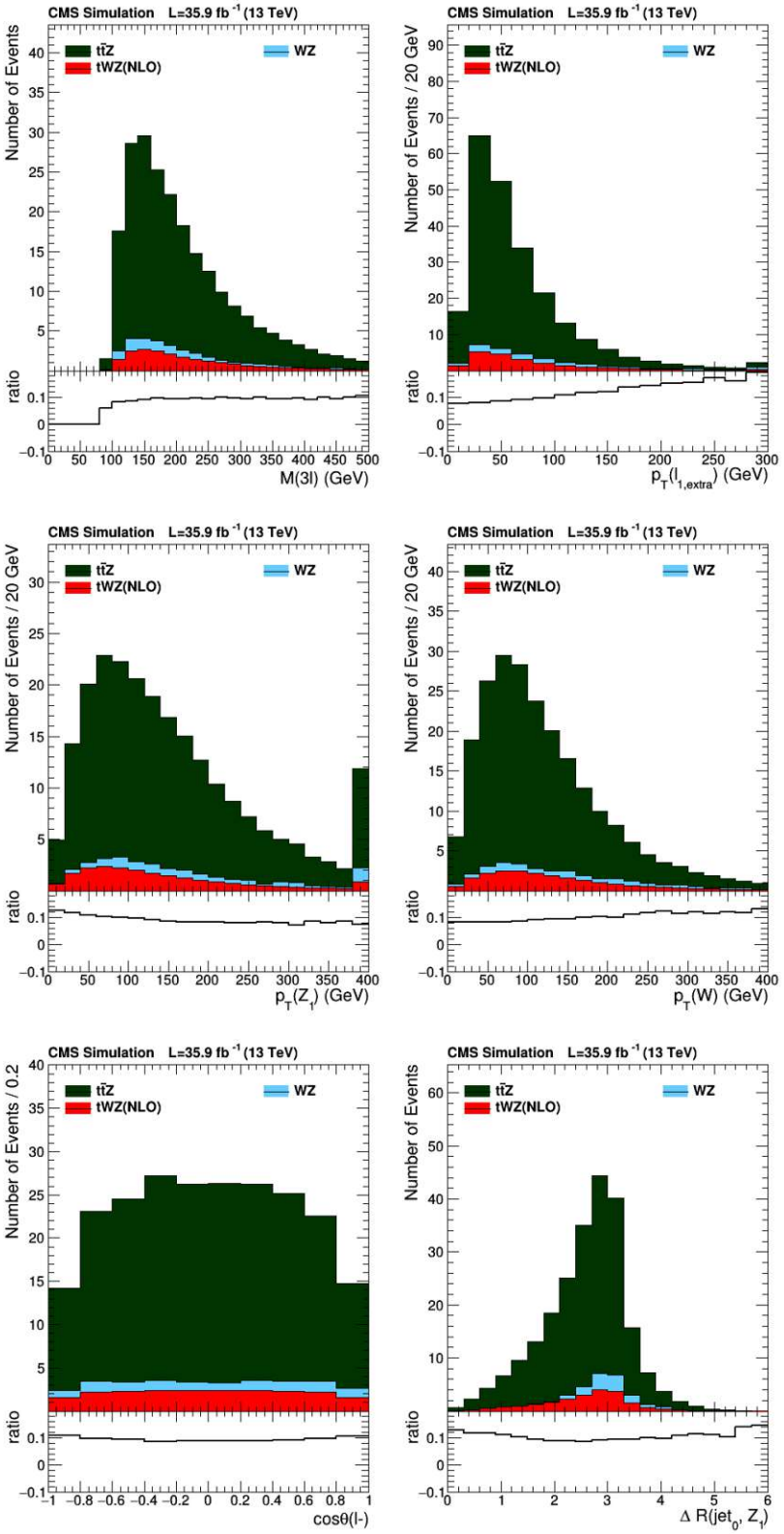

Figure 5.7: TWZ signal region plots for different variables.

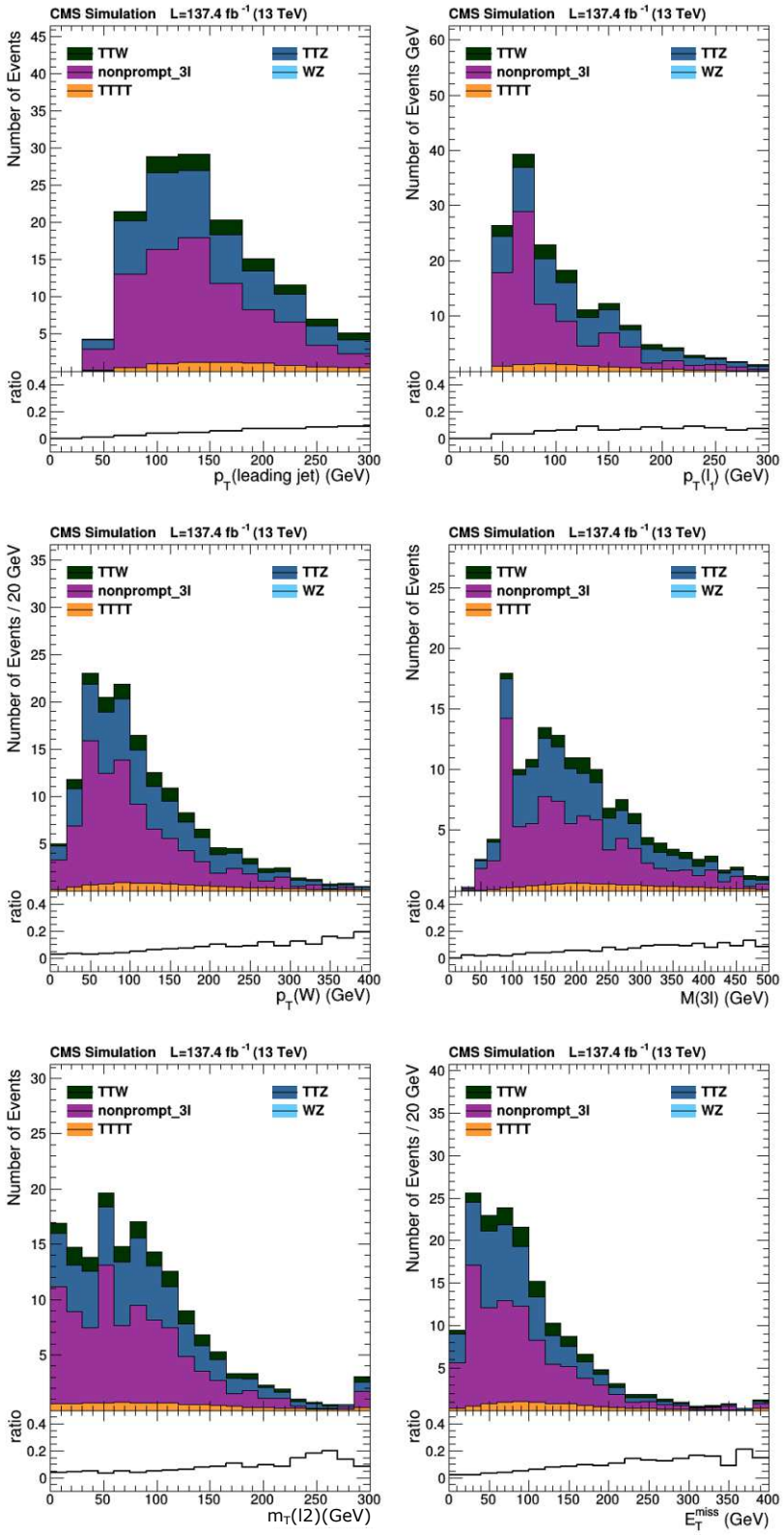

Figure 5.8: 4t signal region plots for different variables.

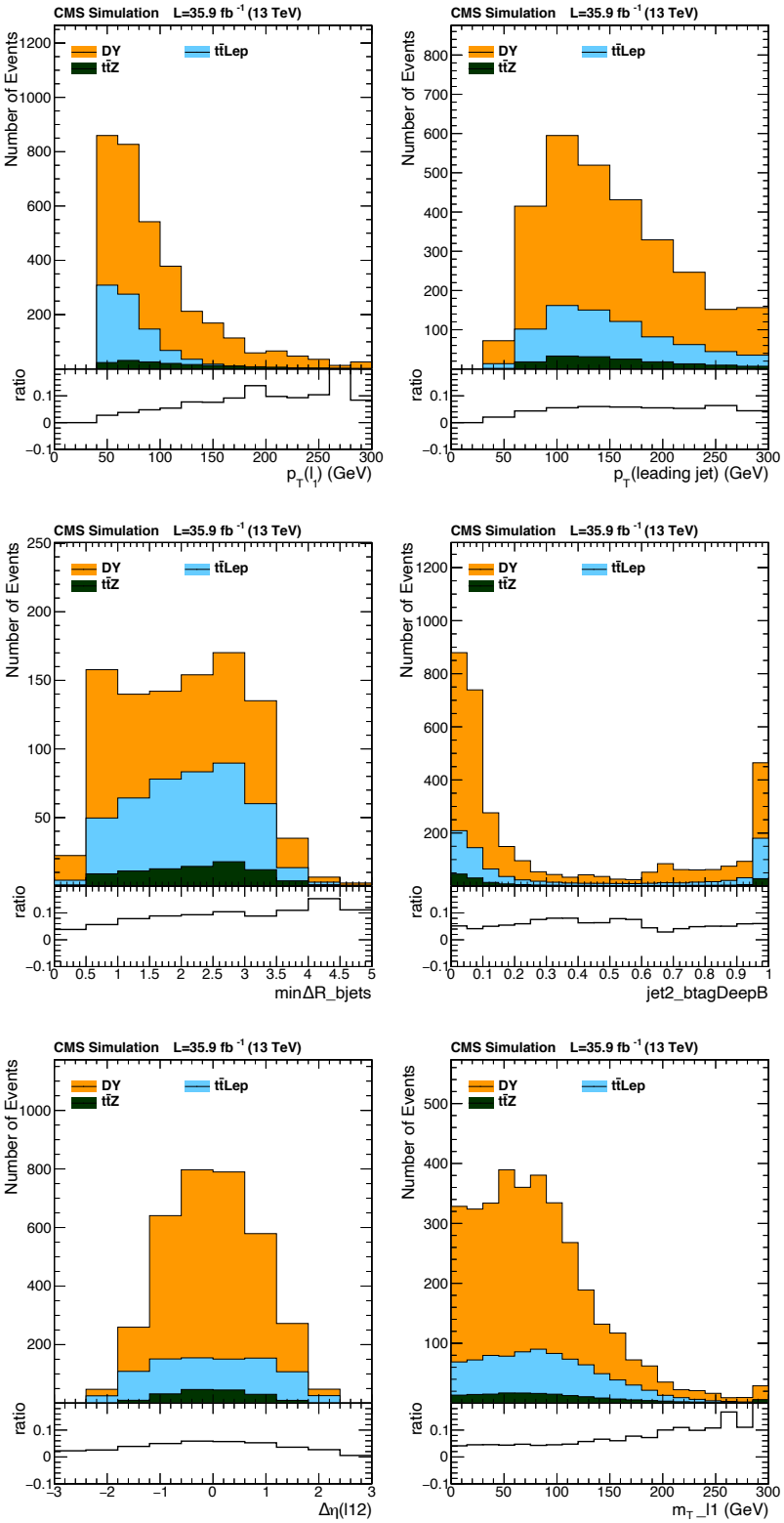

Figure 5.9: $t\bar{t}Z$ signal region plots for different variables.

5.5 Background discrimination

For each process trainings with and without LSTMs were performed. Associated discriminator distributions are shown in Figure 5.10. The usage of LSTMs results in a slightly better discrimination compared to standard machine learning algorithms in all cases. In addition, different settings of ML algorithms were applied. The number of cells in the LSTM layers were varied, plots shown are for 10 cells.

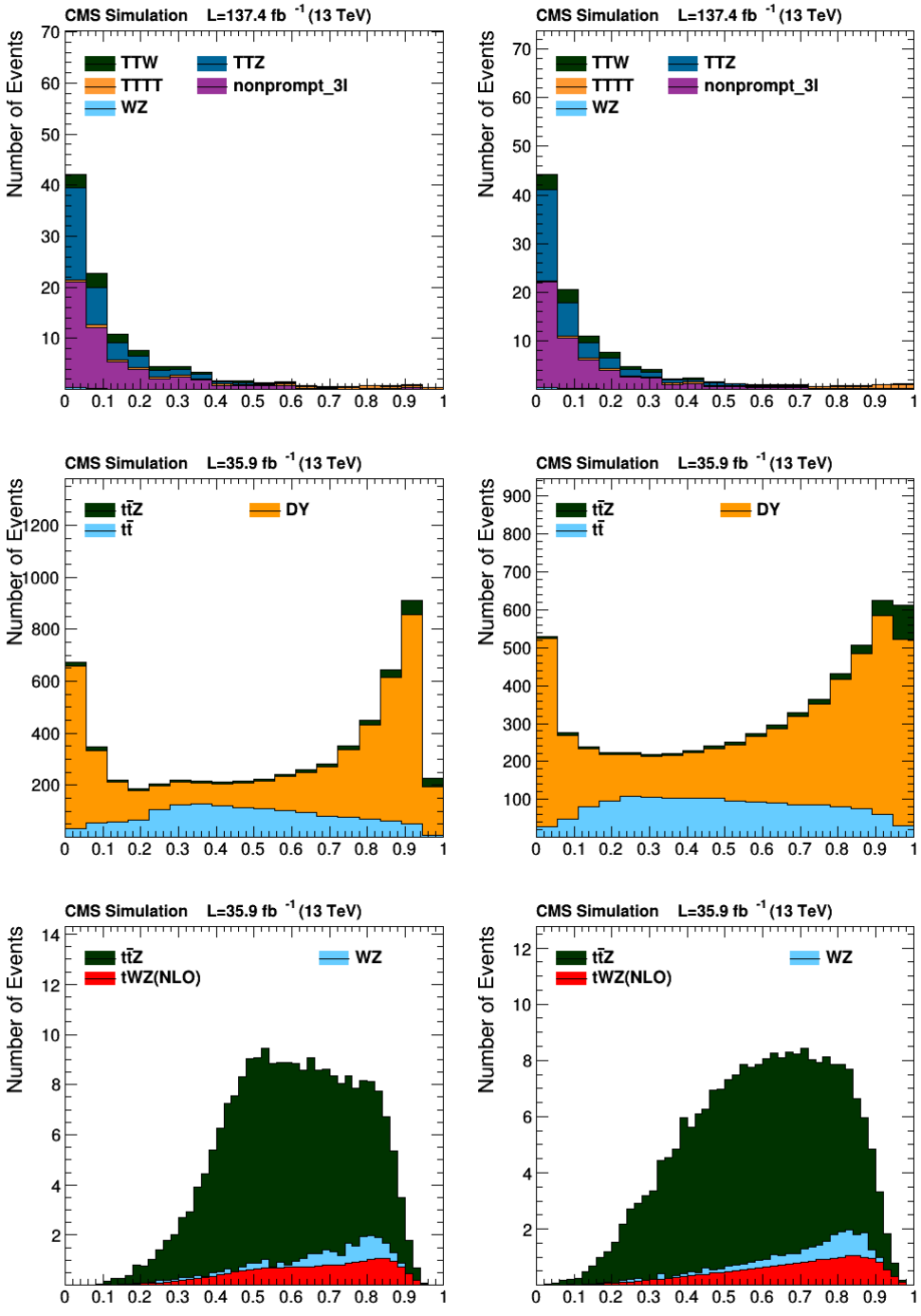

Figure 5.10: The ML discriminator distributions with(right) and without(left) LSTMs for the two different processes: 4t(top), ttZ(middel) and tWZ(bottom).

5.6 Limit calculation

The Higgs Physics Analysis Group of CMS created a tool providing many statistical analysis options [63][64]. The so called combine tool is used for this thesis in order to calculated expected cross-section limits as well as significances, which are described in more detail below. Uncertainties are thereby taken into account and are described in the following section 5.6.2.

5.6.1 Asymptotic Frequentist Limits and Significances

In particle physics it is common to use statistical methods to test hypotheses. In order to explain the concepts used in this thesis some statistical notions are described below closely following [65].

Hypotheses need to be defined to test them. In our case, the null hypothesis H_0 is background only and the alternative hypothesis H_a is signal plus background. The background is constant whereas the signal scales linearly with the signal strength modifier r, which is the main subject of the calculations.

The value q used to perform a hypotheses test is called the test statistics. The Neyman-Pearson $q_N$and the profiled likelihood (PL) q_L are commonly used:

$$q_N = -2\ln\frac{L(H_0)}{L(H_1)} \tag{5.6}$$

where L is the likelihood function

$$L(H) = P(x|H) \tag{5.7}$$

with x being data. The PL depends on the chosen hypothesis and when testing H_0 is

$$q_L = -2\ln\frac{L(b)}{L(\hat{r}s+b)} \tag{5.8}$$

where b stands for background, s for signal and $\hat{r}$ is the maximum likelihood estimator of r.

Assuming the H_0 is correct, the probability of the obtained results being at least as extreme as the observed ones is called the p-value. Here, p_r refers to the signal plus background hypothesis whilst p_b is the p-value of the background only hypothesis. Confidence Levels (CLs) are defined as

$$CLs = \frac{p_r}{1 - p_b} \tag{5.9}$$

and are used to set (upper) limits on a specific parameter, here r is used. An Asimov dataset is one representative set in which all fluctuations are suppressed. In such datasets expected values replace observed quantities.

5.6.2 Uncertainties

Different systematic uncertainties need to be considered for the calculation of cross section limits. In the following, uncertainties used in this thesis are described briefly. In table 5.1 uncertainties used for the three different top quark processes are summariezed.

The MC samples are all normalised to data through the luminosity. The different data taking periods have different luminosities and therefor different uncertainties. In the table the luminositiy uncertainties are listed as "Lumi".

Jet energy scale corrections (JEC) are used to link the measured jet energy deposition to the particle level jet energies [66].

The jet energy resolution (JER) uncertainty considers the fact that the JER in MC jets needs to be smeared to mimic the worse JER in data.

There is an uncertainty on the trigger efficiency which needs to be included, it hast he same value for all processes.

There are b-tag reweighting factors taking into account that the b-tagging algorithm does not perform equally on data and MC. The sacle factors are provided by the CMS BTag group [67]. PDF sets are specified by fitting theoretical predictions to experimental data. Using PDFs yields uncertainties smaller than 1%.

Table 5.1: Uncertainties used for limit calculations are listed for the different processes.

Uncertainties			
	tttt	ttZ	tWZ
JEC	1.09	1.09	1.09
JER	1.01	1.01	1.01
btag_heavy	1.04	1.04	1.04
btag_light	1.04	1.04	1.04
trigger	1.01	1.01	1.01
scale	1.01	1.01	1.01
PDF	1.01	1.01	1.01
Lumi 2016	1.025	1.025	1.025
Lumi 2017/8	1.023	1.023	1.023

6 Limit setting and Results

The tWZ, $t\bar{t}t\bar{t}$ and the $t\bar{t}Z$ processes were investigated. Different ML techniques were applied in order to discriminate signals and backgrounds. ML training results were used to plot discriminator shapes. The obtained histograms were used as inputs for datacards to calculate median expected limits as well as expected significances using the Higgs combine tool as described in section. Signal plus background and background only hypotheses corresponding to each process are used to compute the results.

The median expected limits on the signal strength parameter r are calculated. The null-hypothesis H_0 refers to r=0 and takes only the background. If the total yield, including the signal, match the observation perfectly r equals to 1. The likelihood function described above is constructed using Poisson distribution products:

$$L(n|r,\theta) = \prod_{b=1}^{N_{bins}} P\left(n_b|r\nu_{S,b}\prod_{i=1}^{N_n}\alpha_{i,b}^{\theta_i} + \sum_{i=1}^{N_B}\nu_{B,b}\prod_{k=1}^{N_n}\alpha_{i,k,b}^{\theta_i}\right)\prod_{i=1}^{N_n}N(\theta_i|0,1). \tag{6.1}$$

The total number of bins is N_{bins}, N_n is the number of nuisance parameters θ, N_B the number of backgrounds(B) for each process, α are the used uncertainty values, n_b the yields per bin b and ν_b the number of predicted events in a bin for signal ($\nu_{S,b}$) or backgrounds ($\nu_{B,b}$). Using Eq. 5.8 the profiled likelihood ratio can be written as

$$q(r) = -2ln\frac{L(n|r,\theta)}{L(n|\hat{r},\hat{\theta})} \tag{6.2}$$

where $(\hat{r},\hat{\theta})$ maximize L. The signal strength modifier r is determined by minimizing q(r).

The significance can then be calculated by dividing r by its error: $\frac{r}{\sigma(r)}$. The median expected limits on the signal strength r and the significance are calculated using 95% CL.

For each process calculations were performed for different integrated luminosities, up to 300 fb^{-1}, which is planned to be the total provided by LHC RunIII. Calculation results are shown in Figure 6.1. Each plot states the underlying top quark process including the used object selection.

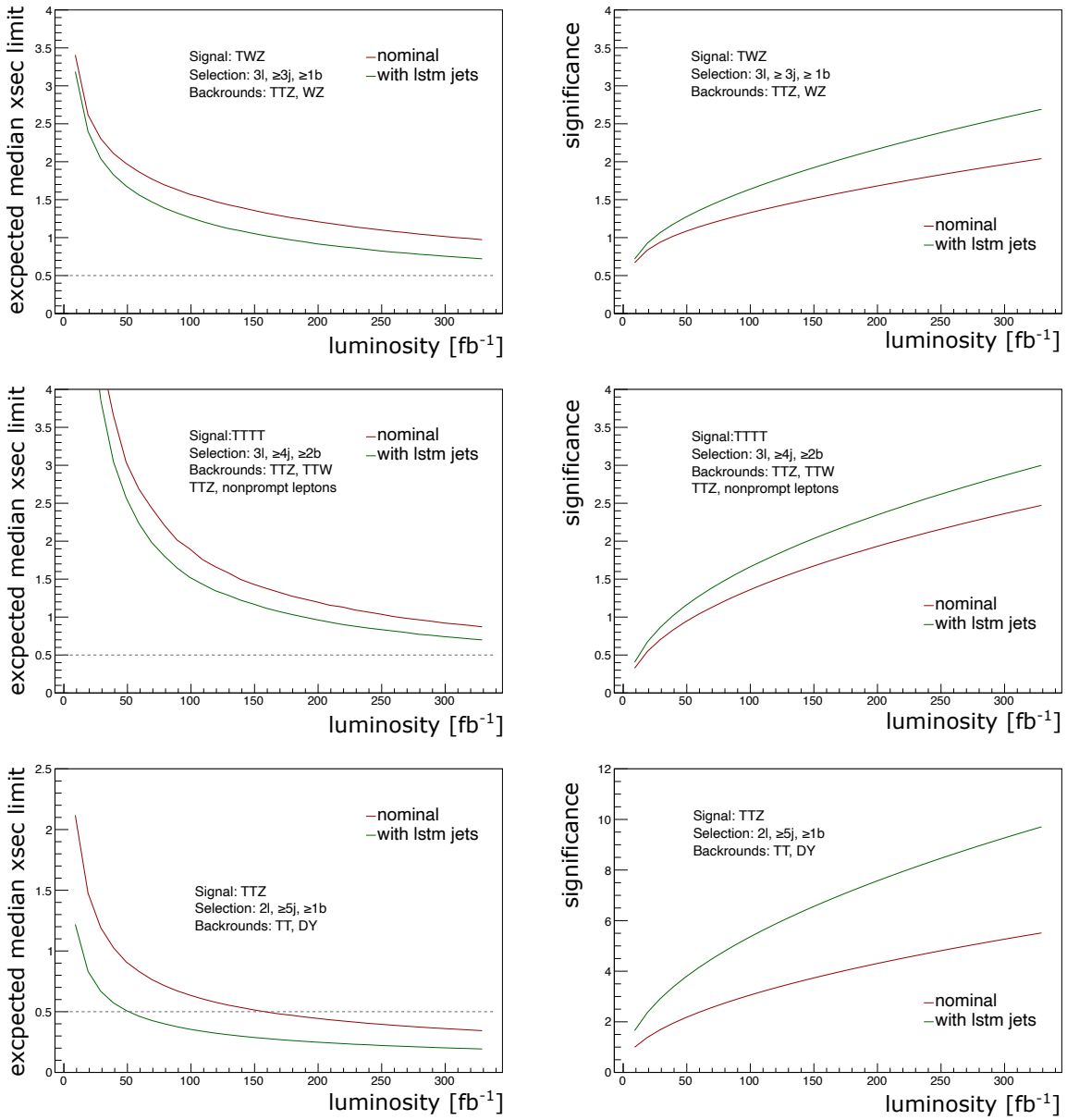

Figure 6.1: Calculated expected cross section limits (right) and significanes (left) for different luminosities are shown for the tWZ (top), the $t\bar{t}t\bar{t}$ (middle) and the $t\bar{t}Z$ (bottom) processes.

7 Conclusion

In the course of this thesis, machine learning techniques were used to improve discrimination of signal and backgrounds. Trainings were performed for three different processes, namely the $t\bar{t}Z$, the tWZ and the $t\bar{t}t\bar{t}$ process.

Machine learning could improve the separation of the signal in all three cases. Furthermore, it was shown that Long Short Term Memories prove to be even more effective in signal-background discrimination. Expected cross section limits as well as significanes for each process were calculated to quantify results and are shown in Figure 6.1. For the reference point of 137.4 fb^{-9}, which corresponds to the luminosity of the LHC Run II, the results for the different processes improved as stated in the following. For the tWZ process the usage of LSTMs decreased the limit form 5.43 to 4.49 and rose the significance from 0.4 to 0.49. The $t\bar{t}t\bar{t}$ limits improved from 1.49 to 1.22; the significane calculated using LSTMs changed from 0.61 to 0.95. The trainings for the $t\bar{t}Z$ process resulted in an improvement of the limit from 0.49 to 0.29 and the significance from 3.6 to 6.32.

The LHC Run III will provide a total integrated luminosity of over 300 fb^{-1}. This amount of data may be enough to discover the tWZ and $t\bar{t}t\bar{t}$ processes and improve measurements of $t\bar{t}Z$. The usage of LSTMs can have a positive impact on the distinction of signals and backgrounds in different physics analyses.

List of Figures

List of Tables

References

[1] M. Thomson, *Modern particle physics*. Cambridge University Press, 2013.

[2] B. Povh, K. Rith, C. Scholz, F. Zetsche, and W. Rodejohann, *Teilchen und Kerne: eine Einführung in die physikalischen Konzepte*. Springer-Verlag, 2013.

[3] S. Weinberg, "A model of leptons," *Physical Review Letters*, vol. 19, no. 21, pp. 1264–1266, Nov. 20, 1967, Publisher: American Physical Society. DOI: 10.1103/PhysRevLett.19.1264.

[4] A. Salam, "Weak and electromagnetic interactions," in *Selected Papers Of Abdus Salam: (With Commentary)*, World Scientific, 1994, pp. 244–254.

[5] S. Node, Go on a particle quest. [Online]. Available: https://sciencenode.org/spotlight/go-particle-quest-first-cern-hackfest.php, accessed: Aug. 2022.

[6] The CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 30–61, Sep. 17, 2012, ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021.

[7] The ATLAS Collaboration, "Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC," *arXiv:1207.7214 [hep-ex]*, Aug. 31, 2012. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214.

[8] W. N. Cottingham and D. A. Greenwood, *An introduction to the standard model of particle physics*. Cambridge university press, 2007.

[9] S. L. Glashow, "Partial-symmetries of weak interactions," *Nuclear physics*, vol. 22,

[10] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 16 Oct. 1964. DOI: 10.1103/PhysRevLett.13.508. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[11] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, 9 Aug. 1964. DOI: 10.1103/PhysRevLett.13.321. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[12] J. Goldstone, A. Salam, and S. Weinberg, "Broken symmetries," *Phys. Rev.*, vol. 127, pp. 965–970, 3 Aug. 1962. DOI: 10.1103/PhysRev.127.965. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.127.965.

[13] H. Schopper, *Particle Physics Reference Library: Volume 1: Theory and Experiments*. Springer Nature, 2020.

[14] N. K. Rad, "Search for supersymmetric partners of the top quark in models with compressed mass spectra with the cms detector," Ph.D. dissertation, Vienna, OAW, 2018.

[15] C. Campagnari and M. Franklin, "The discovery of the top quark," *Reviews of Modern Physics*, vol. 69, no. 1, p. 137, 1997.

[16] K. Hagiwara, K. Hikasa, K. Nakamura, M. Tanabashi, M. Aguilar-Benitez, C. Amsler, R. Barnett, P. Burchat, C. Carone, C. Caso, *et al.*, "Review of particle physics," *Physical Review D (Particles and Fields)*, 2002.

[17] M. Waqas, "Search for the associated production of a Z boson with a single top quark using CMS data at 8 TeV and 13 TeV and performance studies of the CMS silicon tracker.," Presented 04 Sep 2019, Feb. 2019. [Online]. Available: https://cds.cern.ch/record/2779481.

[18] J. h. Kuhn, A. Scharf, and P. Uwer, "Weak interactions in top-quark pair production at hadron colliders: An update," *Physical Review D*, vol. 91, p. 014 020, 2015.

[19] N. Poudyal, "Measurement of the cross section of top quark pairs produced in association with a photon in lepton+ jets events at *sqrts*= 13 tev with full run ii cms data," Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States), Tech. Rep., 2021.

[20] C. H. James, "Branching ratio," in *Encyclopedia of Astrobiology*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 218–218, ISBN: 978-3-642-11274-4. DOI: 10.1007/978-3-642-11274-4_206. [Online]. Available: https://doi.org/10.1007/978-3-642-11274-4_206.

[21] P. Zyla *et al.*, "Review of Particle Physics," *PTEP*, vol. 2020, no. 8, p. 083C01, 2020. DOI: 10.1093/ptep/ptaa104.

[22] S. Abe, T. Ebihara, S. Enomoto, K. Furuno, Y. Gando, K. Ichimura, H. Ikeda, K. Inoue, Y. Kibe, Y. Kishimoto, *et al.*, "Precision measurement of neutrino oscillation parameters with kamland," *Physical Review Letters*, 2008.

[23] R. N. Mohapatra, "Seesaw mechanism and its implications," in *SEESAW 25*, World Scientific, 2005, pp. 29–44.

[24] D. Spitzbart, "Search for supersymmetric partners and anomalous couplings of the top quark with the cms experiment," Ph.D. dissertation, Wien, 2019.

[25] K. Arun, S. B. Gudennavar, and C. Sivaram, "Dark matter, dark energy, and alternate models: A review," Jul. 2017, Accepted: 2020-11-11T01:38:52Z Publisher: Elsevier B.V, ISSN: 0273-1177.

[26] L. Evans and P. Bryant, "Lhc machine," *Journal of instrumentation*, vol. 3, no. 08, S08001, 2008.

[27] M. e. a. Banner, "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the cern pp collider," *Physics Letters B*, vol. 122, no. 5, pp. 476–485, 1983, ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91605-2. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0370269383916052.

[28] G. e. a. Arnison, "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 gev," *Physics Letters B*, vol. 122, no. 1, pp. 103–116, 1983, ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91177-2. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0370269383911772.

[29] L. Di Lella and C. Rubbia, "The discovery of the w and z particles," in *60 Years of CERN Experiments and Discoveries*, World Scientific, 2015, pp. 137–163.

[30] F. Marcastel, "Cern's accelerator complex. la chaıne des accélérateurs du cern," *Στo:(Oκτ. 2013). General Photo*, 2013.

[31] A. Goshaw, "The atlas experiment at the cern large hadron collider [j]," *Center for Strategic and International Studies*, 2009.

[32] The ALICE Collaboration, "The alice experiment at the cern lhc," *Jinst*, vol. 3, no. 420, S08002, 2008.

[33] C. Collaboration, S. Chatrchyan, G. Hmayakyan, V. Khachatryan, A. Sirunyan, W. Adam, T. Bauer, T. Bergauer, H. Bergauer, M. Dragicevic, *et al.*, "The cms experiment at the cern lhc," *JInst*, vol. 3, S08004, 2008.

[34] The LHCb Collaboration et. al., "The lhcb detector at the lhc," *Journal of instrumentation*, vol. 3, no. 08, S08005–S08005, 2008.

[35] W. Herr and B. Muratori, "Concept of luminosity," 2006.

[36] The CMS Collaboration, "The CMS experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 8, S08004, Aug. 14, 2008, Publisher: IOP Publishing, ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/S08004.

[37] ——, Detector. [Online]. Available: https://cms.cern/detector, accessed: Sept. 2020.

[38] D.Barney, CMS Detector Slice. [Online]. Available: https://cds.cern.ch/record/2120661, accessed: Okt. 2020.

[39] University of Zurich, Simple example of 3D axes with spherical coordinates. [Online]. Available: https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates, accessed: Okt. 2020.

[40] The CMS Collaboration, Solenoid. [Online]. Available: https://cms.cern/detector/bending-particles, accessed: Okt. 2020.

[41] ——, "The CMS Trigger System," *arXiv:1609.02366 [hep-ex, physics:physics]*, Jan. 25, 2017. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366.

[42] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, "Machine learning at the energy and intensity frontiers of particle physics," *Nature*, 2018.

[43] M. Feickert and B. Nachman, "A living review of machine learning for particle physics," *arXiv preprint arXiv:2102.02770*, 2021.

[44] D. Bourilkov, "Machine and deep learning applications in particle physics," *International Journal of Modern Physics A*, vol. 34, no. 35, p. 1 930 019, 2019.

[45] A. Hocker, X. Prudent, J. Therhaag, Y. Mahalalel, M. Backes, R. Ospanov, M. Kruk, M. Jachowski, A. Voight, A. Robert, *et al.*, "Tmva-toolkit for multivariate data analysis with root: Users guide," Tech. Rep., 2007.

[46] T. L. Brückler, "Deep learning in searches for supersymmetry with compressed mass spectra," Ph.D. dissertation, Wien, 2020.

[47] S. Hayas, *Neural Networks and Learning Machines*. CPearson Endcation, Inc., 2009.

[48] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[49] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.

[50] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

[51] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[52] L. Arras, J. Arjona-Medina, M. Widrich, G. Montavon, M. Gillhofer, K.-R. Müller, S. Hochreiter, and W. Samek, "Explaining and interpreting lstms," in *Explainable ai: Interpreting, explaining and visualizing deep learning*, Springer, 2019, pp. 211–238.

[53] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *CoRR*, vol. abs/1808.03314, 2018. arXiv: 1808.03314. [Online]. Available: http://arxiv.org/abs/1808.03314.

[54] S.-S. Baek, J. Pyo, and J. A. Chun, "Prediction of water level and water quality using a cnn-lstm combined deep learning approach," *Water*, vol. 12, no. 12, 2020, ISSN: 2073-4441. [Online]. Available: https://www.mdpi.com/2073-4441/12/12/3399.

[55] J. Schmidhuber, Our Decade of DeepLearning. [Online]. Available: https://people.idsia.ch/~juergen/2010s-our-decade-of-deep-learning.html, accessed: Aug. 2022.

[56] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, pp. 493–498, 2019.

[57] F. Chollet *et al.*, *Keras*, https://keras.io, 2015.

[58] T. Tschida, "Cross section measurements of processes with a single top quark and two vector bosons with the cms experiment," Ph.D. dissertation, Wien, 2020.

[59] R. Frederix, D. Pagani, and M. Zaro, "Large nlo corrections in ttw and tttt hadroproduction from supposedly subleading ew contributions," *Journal of High Energy Physics*, vol. 2018, no. 2, Feb. 2018, ISSN: 1029-8479. DOI: 10.1007/jhep02(2018)031. [Online]. Available: http://dx.doi.org/10.1007/JHEP02(2018)031.

[60] J. M. Campbell and R. K. Ellis, "Ttw production and decay at nlo," *Journal of High Energy Physics*, vol. 2012, no. 7, pp. 1–12, 2012.

[61] The CMS Collaboration et al., "Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Journal of High Energy Physics*, vol. 2020, no. 3, p. 56, Mar. 10, 2020, ISSN: 1029-8479. DOI: 10.1007/JHEP03(2020)056.

[62] S. M. A. Mortier, "Exploring top-quark and z boson coupling in top pair production in association with a z boson," Ph.D. dissertation, Gent U., 2020.

[63] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C*, vol. 71, no. 2, pp. 1–19, 2011.

[64] The CMS Collaboration, Higgs Analysis-Combined Limit. [Online]. Available: http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/, accessed: Sept. 2020.

[65] E. Gross, "Practical statistics for high energy physics," *CERN yellow reports: school proceedings*, vol. 4, no. 0, p. 165, 2017.

[66] The CMS Collaboration, Jet Energy Corrections. [Online]. Available: https://twiki.cern.ch/twiki/bin/viewauth/CMS/IntroToJEC, accessed: Sept. 2020.

[67] ——, bTag Scale Factor Methods. [Online]. Available: https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods, accessed: Sept. 2020.

Appendices

Table 1: All variables used for training purposes of the three different top quark processes are listed in this table.

	ttZ	tWZ	tttt
global event properties	nJetGood	ht	nJetGood
	nBTag	met_pt	nBTag
	met_pt	m3l	nlep
	ht	nJetGood	ht
		nBTag	htb
			ht_ratio
			met_pt
Z	mll		
jet kinematics	jet0_pt	jet0_pt	jet0_pt
	jet0_eta	jet0_eta	jet0_eta
	jet0_btagDeepB	jet0_btagDeepB	jet0_btagDeepFlavB
	jet1_pt	jet1_pt	jet1_pt
	jet1_eta	jet1_eta	jet1_eta
	jet2_pt	jet1_btagDeepB	jet1_btagDeepFlavB
	jet2_eta	jet2_pt	jet2_pt
	jet3_pt	jet2_eta	jet2_eta
	jet4_pt	jet2_btagDeepB	jet2_btagDeepFlavB
	jet5_pt		jet3_pt
			jet4_pt
			jet5_pt
			jet6_pt
			jet7_pt
			dEtaj_12
			dPhij_12
			mj_12
lepton kinematics	l1_pt	lnonZ1_pt	l1_pt
	l1_eta	lnonZ1_eta	l1_eta
	l2_pt		l2_pt
	l2_eta		l2_eta
	l1_phi		dEtal_12

	l2_phi	dPhil_12
	l1_relIso	mT_l1
	l2_relIso	mT_l2
	minDR_jl1	ml_12
	minDR_bjl1	
	minDR_jl2	
	minDR_bjl1	
lepton/jet		mlj_11
		mlj_12
#leptonMVA	l1_mvaTOP	
	l2_mvaTOP	
	l3_mvaTOP	
	l1_mvaTOPWP	
	l2_mvaTOPWP	
	l3_mvaTOPWP	
nonZ1_l1 vs. other objects	jet0_nonZl1_deltaR	
	jet1_nonZl1_deltaR	
	bJet_Z1_deltaR	
	bJet_non_Z1l1_deltaR	
	maxAbsEta_of_pt30jets	
Z1 vs. other objects	nonZ1_l1_Z1_deltaPhi	
	nonZ1_l1_Z1_deltaR	
	jet0_Z1_deltaR	
	jet1_Z1_deltaR	
	jet2_Z1_deltaR	
Z1 kinematics	Z1_pt	
	Z1_eta	
	Z1_cosThetaStar	
leptonic W	W_pt	