Phenomenology of $\mathrm{U}(1)_{L_\mu - L_\tau}$ Gauge Extension of the Standard Model at LHC

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Meng Shi
aus
Anhui, China

Bonn, January 2021

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter:	Prof. Dr. Manuel Drees
2. Gutachter:	Prof. Dr. Herbert Dreiner
Tag der Promotion:	20.04.2021
Erscheinungsjahr:	2021

Abstract

Extending the Standard Model (SM) by a $U(1)_{L_\mu - L_\tau}$ group gives potentially significant new contributions to $g_\mu - 2$, allows the construction of realistic neutrino mass matrices, incorporates violation of lepton universality violation, and offers an anomaly–free mediator for a Dark Matter (DM) sector. In this thesis, we focus on constraints on this model from the Large Hadron Collider (LHC) data based on final states with two, three, or four leptons and missing energy. On the one hand, We recast a large number of LHC analyses involving relevant final states using the `CheckMATE` framework; and set the upper LHC limit. However, most of the existing LHC analyses are not specially designed for this model, which leads to a possibility of optimizing the cuts. Therefore, on the other hand, we apply several Machine Learning (ML) algorithms to distinguish the model from the SM using simulated LHC data; and derive the bound accordingly. We also try to uncover the "black box" of ML by discussing the feature importance, which is extracted from Gradient Boosting Decision Tree (GBDT). The same scheme could be generalized to other models.

Acknowledgements

I would like to thank many people who helped me during my PhD study in Bonn. As a foreigner and chemistry graduates, it is not easy for me in both personal and academic life in the past years.

My greatest gratitude goes to my supervisor Manuel Drees for his help and support of my study and research. Not only his incisive insights in physics but also his enriched knowledge of many other things have inspired me a lot. I am also very thankful to Prof. Herbert Dreiner to read and referee my thesis, and to Prof. Ian Brock and Prof. Helmut Baltruschat for joining my doctoral committee.

I am grateful to my current and former colleagues at BCTP: Zhongyi Zhang, Yong Xu, Zeren Simon Wang, Bardia Najjari Farizhendi, Rahul Mehra, Lina, Ernany Rossi Schmitz, Fazlollah Hajkarim, etc., for all the happy conversations. In particular, I would like to thank Yong and Zhongyi, for reading my thesis, and for all the help and support.

I would also like to give my gratitude to the secretaries: Petra Weiss, Patricia Zündorf, Christa Börsch, Andreas Wißkirchen, especially for Andreas Wißkirchen's kind and patient help.

Finally, I should thank my grandfather, my parents, my brother and my wife, for their love and support in my life.

List of Publications

The thesis contains results that have been published in

1. M. Drees, M. Shi and Z. Zhang,
   Constraints on $U(1)_{L_\mu - L_\tau}$ from LHC Data,
   arXiv: 1811.12446 [hep-ph]

and the work will be published soon

2. M. Drees, M. Shi and Z. Zhang,
   Machine Learning Optimized Search for $U(1)_{L_\mu - L_\tau}$ at LHC

Contents

CHAPTER 1

Introduction

1.1 The Standard Model and Beyond

The Standard Model (SM) is a theory which describes electroweak and strong interactions in the Universe. It has been widely tested in numerous experiments like Large Hadron Collider (LHC) and shown great success in the last decades. Among them the electron anomalous magnetic moment measurement [1, 2] is one of the most precise tests within the SM; its prediction agrees with experimental result with more than 10 significant figures.

The SM is a quantum field theory respecting to gauge symmetry and Poincaré symmetry. Its fundamental constituents are shown in Fig. 1.1. They can be divided into two groups. One major group is the fermions including quarks and leptons, and they form the normal matter in our Universe. From the point of view of group theory, fermions are the irreducible representations of Poincaré group [3], and there are three generations. Each generation of quarks or leptons has a pair of fermions, like (u and d). Each cell of fermions in Fig. 1.1 is called a flavor, studying the behavior of flavor changing is a hot topic in particle physics, which is the flavor physics [4, 5].

Another group is the bosons, which are the force carriers between fermions. The representations of the SM gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$. $SU(3)_C$ mediates the strong interactions between quarks, and the corresponding gauge bosons are massless gluons. The subscript C represents colors, and there are in total three different colors (red, green and blue) for each type of gluons and quarks. $SU(2)_L \times U(1)_Y$ mediates the electroweak interactions, a universal description of electromagnetic and weak interaction at high energy level. It is also known as the Weinberg-Salam theory, proposed by Sheldon Glashow [6], Abdus Salam [7] and Steven Weinberg [8]. The subscript L represents left-handed fermions, which means only left-handed fermions are involved, therefore the parity is not conserved under electroweak interaction. The corresponding gauge bosons are called W^+, W^- and Z bosons, and they are massive.

The last piece is the Higgs boson, the only spin-0 particle in the SM. It is a core part of the SM, and massive bosons acquires their masses through Higgs Mechanism. During the so-called Electroweak Symmetry Breaking (EWSB) process, the SM gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$ breaks into $SU(3)_C \times U(1)_{EM}$. Before EWSB, if we only consider EW bosons, namely the massless W and Z bosons and the Higgs doublet, we have in total $3 \times 2 + 4 = 10$ degrees of freedom (d.o.f). After EWSB, W and Z bosons acquire their masses by eating 3 d.o.f of Higgs doublet, the left one is the Higgs boson h, which was discovered in 2012 [9, 10]. Then we still have $3 \times 3 + 1 = 10$ d.o.f. left. As for

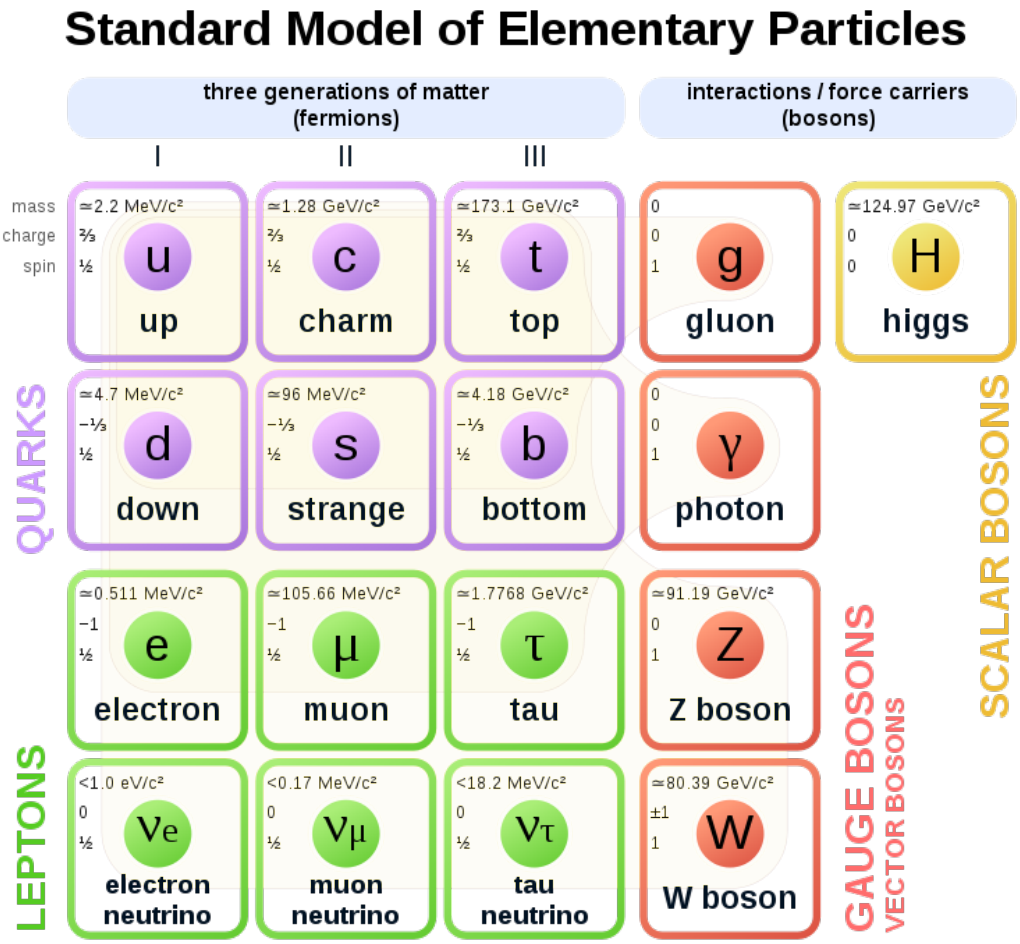

Figure 1.1: Elementary particles of the Standard Model from Wikipedia.

fermions, they are also massless before EWSB, their masses are obtained by interacting with Higgs field, and proportional to the corresponding Yukawa couplings.

The simplified SM Lagrangian is as follows [11],

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\Psi}\not{D}\Psi + \bar{\Psi}_i y_{ij}\Psi_j H + |D_\mu H|^2 - V(H) + h.c. \tag{1.1}$$

the first term $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ is the product of field strength tensor, and it contains interactions between groups of gauge bosons. The second term $i\bar{\Psi}\not{D}\Psi$ describes fermion interactions mediated by gauge bosons, where D is the covariant derivative containing gauge fields. The third term $y_{ij}\bar{\Psi}_i\Psi_j H$ is the Yukawa interaction term, where y_{ij} is the Yukawa couplings. Finally, $V(H)$ is the scalar potential of Higgs field, which is responsible for EWSB.

Although the SM is quite successful, and its parameters are being precisely measured by experiments. There are still a lot of questions that the SM can not offer satisfied answers. To be more specific, we

first summarize Feynman diagrams involving fermions in the SM, as shown in Fig. 1.2.

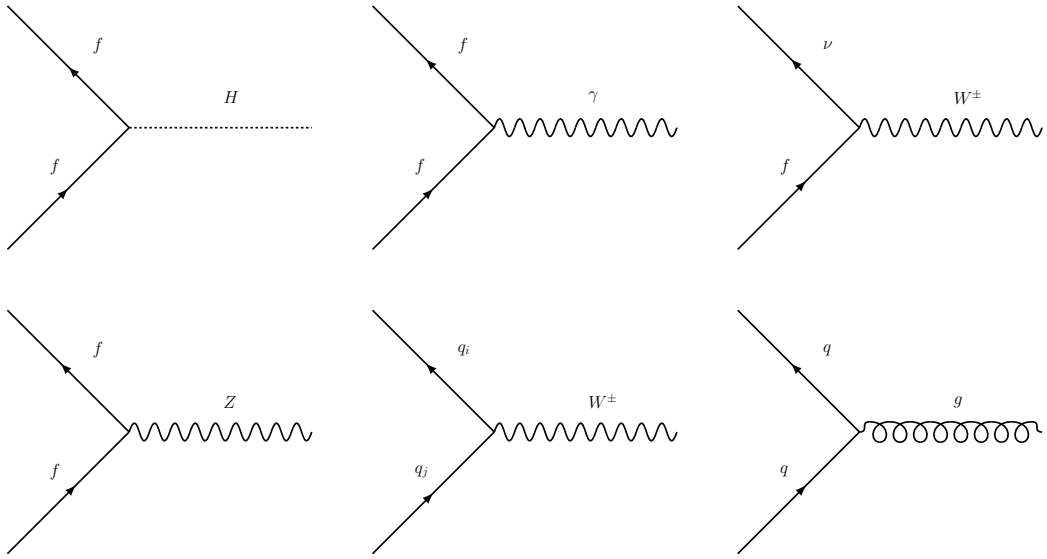

Figure 1.2: The SM Feynman diagrams involving fermions. Here f is a fermion, l is charged lepton, and q is quark. The plot is from a talk "Flavor Physics" at Terascale Summer School 2020 [12].

- Matter-antimatter Asymmetry: It is well known that our Universe is dominated by matter, and the observed antimatter is quite few; otherwise they will annihilate with matter, which makes our universe nothing interesting but only leftover energy. However, from Fig. 1.2, we can see that, antimatter always pair up with matter, which means that at the beginning of our Universe, when the Big Bang happens, the number of matter and antimatter should be equal in SM. During the process of Universe expanding, if they respect the SM, the left matter and antimatter should remain equal in our current Universe. But this is clearly not the case. So it is widely believed that resolving the matter and antimatter asymmetry problem needs physics beyond the SM [13].

- B Physics Anomaly: In the SM, as shown in Fig. 1.2, quark can only change its flavor through $W+$ or $W-$ boson. Moreover, it can only be changed from up-type to down-type at tree level, or vice versa, which is suppressed by Cabibbo–Kobayashi–Maskawa (CKM) matrix. The same type flavor changing, like $b \to s$, is induced at loop level. A measurement from LHCb [14] shows that, the differential branching fraction of process is $\mathcal{B}(B_0 \to \phi\mu^+\mu^-) = (7.97^{+0.45}_{-0.43}) \times 10^{-7}$, which is more than 3σ below the SM prediction. This anomaly implies there might be extra contributions to this process.

- Lepton Universality Violation: In SM, the charged leptons (electrons, muons and taus) share the same coupling in weak interaction. Therefore the process of $B^+ \to K^+\mu^+\mu^-$ should have the same cross section with process of $B^+ \to K^+e^+e^-$ in SM. However, LHCb [15] found that, the ratio of these two branching fractions is $R_K = 0.846^{+0.060}_{-0.054}$. It is 2.5σ deviation from the SM prediction 1, which is also a hint of a BSM theory.

- Hierarchy Problem: As we have mentioned earlier, fermions (except neutrinos) gain their masses through Yukawa interaction term with masses proportional to Yukawa couplings. But, if we

check their observed masses in Fig. 1.1, we find that they are incredibly different. For example, the mass of the heaviest quark M_t is nearly 5 order of magnitude larger than the lightest one M_u. Even for the same generation, M_t is still 2 order of magnitude heavier than M_b. SM cannot explain this discrepancy, since they all originate from the same symmetry breaking and proportional to Yukawa couplings within SM.

- Higgs Mass: The SM is a renormalizable theory, and observables at a low energy level would get quantum corrections from high energy cutoff scale. In order to get a well-defined quantity, the infinite correction terms should cancel each other. In SM, the higgs mass M_h term gets fermion loop corrections (such as top-quark loop correction), but these infinities cannot be canceled within SM, which leads to a higgs mass proportional to high energy cutoff scale Λ, like Planck scale $M_P \approx 10^{19}$ GeV. This is clearly not the case observed in experiments, since the measured higgs mass is around 125 GeV [9, 10]. One possible explanation is from the well-known supersymmetry (SUSY) theory [16, 17], which is inspired by that, the sign of fermion loop correction and scalar loop correction differs. Then, there is a chance that these corrections cancel each other. Hence SUSY proposed a new set of particles, named superpartners, to solve this problem.

- Neutrino Masses: In SM, neutrino masses are set to zero according to experimental observations in earlier days, so there is no neutrino flavor mixing. However, neutrino oscillation [18] was observed and awarded as the 2015 Nobel Physics Price. To explain it, a naive idea is adding three right handed neutrinos into SM, then they get their masses by interacting with higgs field, similar as quarks. But, since right handed neutrinos are not observed so far, they must be very heavy. This simple idea cannot explain this mass discrepancy. Another solution is the so-called Seesaw mechanism [19], similar with the naive idea, it introduces three right handed neutrinos, but here they have Majorana mass terms. In this way, the left handed neutrino mass is proportional to inverse of right handed neutrino mass, so a very large right handed neutrino mass naturally generates a tiny left handed neutrino mass. Moreover, since Majorana fermion is its own anti-particle, there exists clearly lepton number violation in this theory, which indicates BSM theory. It could be tested by the neutrinoless double beta decay $0\nu\beta\beta$, but currently no evidence was found in experiments [20].

- Dark Matter: If we call the matter that makes up all stars and things we “see" the normal matter, there are strong evidences that another type of matter which we cannot “see" exists, called dark matter. It is “dark" since it does not interact with the electromagnetic force. Surprisingly it contributes 27% contents of the Universe; in comparison, normal matter only contributes 5%. In SM, the naive candidate is neutrinos, since they are neutral. But it is well studied, given the masses of neutrinos, they can only attribute to a very small potion of dark matter [21, 22]. Therefore we need to go beyond the SM. Dark matter is one of the main topics in this thesis, we will give a more detailed review in the next chapter.

- Gravity: Although the SM makes a great success in explaining electroweak and strong interaction, it is incapable of embracing gravity into its paradigm. If we quantize the gravity field via the same way as fields in the SM, a dimensional coupling constant G (Newton’s gravitational constant) comes in. This leads a non-renormalizable theory. Fortunately, the gravitational force is even far weaker than electroweak force, it only needs to be considered when the energy scale

is reaching the Planck scale. At current state of experimental energy level, such as 14 TeV at LHC, the gravity is negligible. The SM is believed to be an effective approximating theory to a high energy complete theory, and one of the best candidates is String Theory.

1.2 Collider Physics

Given all these unsolved questions mentioned above, many experiments are designed to give more detailed studies to both SM and BSM processes. In this thesis, we will focus on Large Hadron Collider (LHC) experiments, especially ATLAS (shown in Fig .1.3) and CMS.

LHC is the world's largest and most powerful particle accelerator, it is 27-kilometer circumference long. With this large size, it can accelerate particles to nearly speed of light. LHC is designed to give a precise measurement of the SM phenomenon, like Higgs properties, also to probe new physics beyond the SM. So far, the most successful outcome of LHC is the discovery of Higgs at 2012 [9, 10].

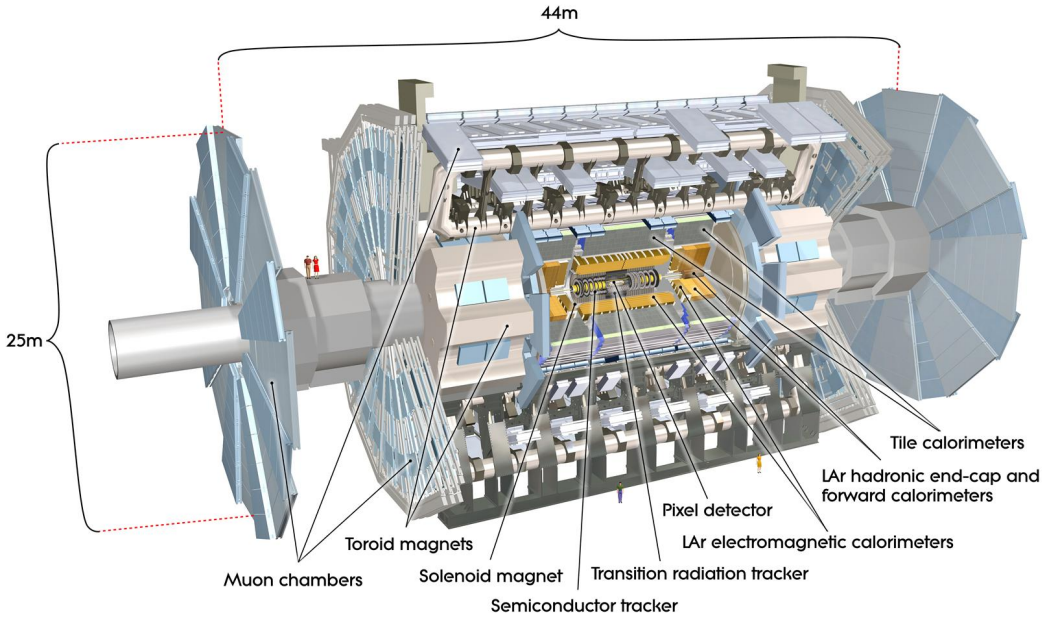

Figure 1.3: The ATLAS detector on ATLAS@CERN website.

The CMS and ATLAS detector is depicted in Fig. 1.4, and it consists of four main parts, which are the inner tracks, electromagnetic calorimeter, hadron calorimeter and the outmost muon chambers. While the beams of particles are accelerated to certain energy, they make collisions and innumerous particles are produced. These particles are identified and measured in the following ways.

- Photons: photons are neutral, so they will leave straight tracks in the inner track detector, and deposited at electromagnetic calorimeter.
- Electrons: Electrons are bended in the inner track detector, and their momenta are measured at this detector. Then, similar to photons, they are also deposited at electromagnetic calorimeter,

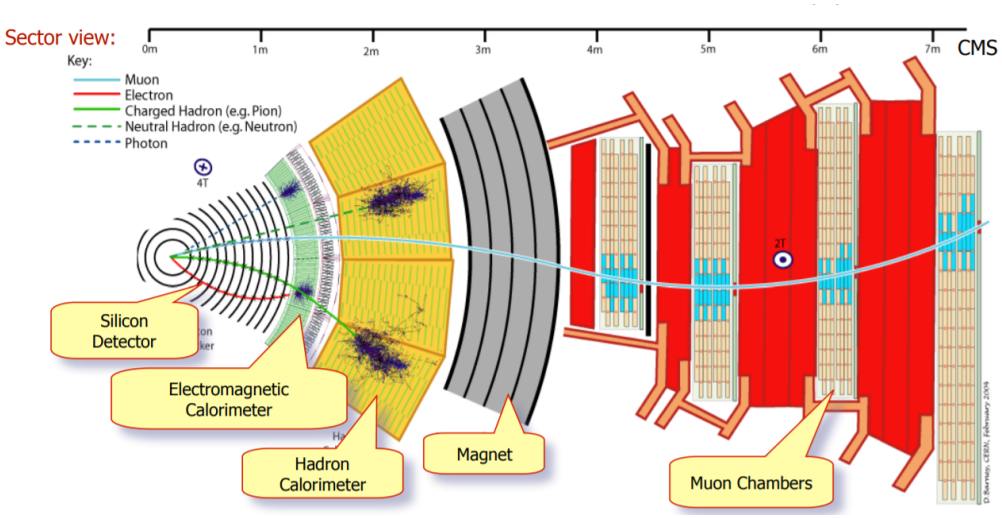

Figure 1.4: The detector section on CMS@CERN website.

with their energy recorded.

- Jets: Jets are reconstructed by their tracks left in the inner tracks, the electromagnetic calorimeter, and the hadron calorimeter using anti-k_T algorithm. Finally, they are deposited at the hadron calorimeter.
- Muons: Muons have similar properties with electrons, except that they are 200 times heavier. With this large mass, muon will not be stopped at the electromagnetic calorimeter. Instead, they will be detected by the outmost muon chambers.
- Missing particles: There are certain particles which cannot be seen in CMS or ATLAS detectors, such as neutrinos, and dark matter. If these particles are stable, they will leave no track in detectors. The only way to see them is through the missing transverse momentum. Since the beams of particles are colliding collinearly, the initial transverse momentum should be zero. Then the final missing transverse momentum is responsible for these undetected missing particles.

So far, the Run 2 of ATLAS and CMS has reached a energy level of $\sqrt{s} = 13$ TeV. With such a high energy, they are appealing to test some BSM theories with high energy particles predicted, such as SUSY. Meanwhile, they have a very high integrated luminosity up to 139 fb^{-1}. This means that, there are more collisions happened and recorded. Then it allows to give a more detailed analysis to some rare processes, also be more confident to some new physics signals. However, LHC also has its cons. The collisions are protons to protons, so the products are like a soup of gluons and different quarks. These highly impure collisions make it difficult to identify and measure produced particles.

1.3 $U(1)_{L_\mu - L_\tau}$ Model

A simple theory which goes beyond the SM is $U(1)_{L_\mu - L_\tau}$ model, which extends the SM gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$ to $SU(3)_C \times SU(2)_L \times U(1)_Y \times U(1)_{L_\mu - L_\tau}$, and is naturally anomaly free.

Unlike other $U(1)$ gauge extension models like $U(1)_{B-L}$, $U(1)_{L_e-L_\mu}$ and $U(1)_{L_e-L_\tau}$, it has no extra coupling with electrons, which allows it evading from most leptonic experimental constraints, such as Large Electron-Position Collider (LEP). [1] A complex scalar field ϕ_{DM} can also be incorporated into this model with a non-trivial charge under the new gauge group; in addition, by setting it odd under a $\mathbb{Z}_2$ symmetry makes it a stable dark matter candidate.

Experimental searches for this model include final states with 2-leptons, 3-leptons, 4-leptons and invisible particles. These relevant events are abundant in LHC, and many of them are unexplored. This makes LHC data a valuable resource to constrain the model. To set the upper limits, one option is to use the `CheckMATE` framework [23, 24] to recast existing LHC analyses. It takes Monte Carlo events generated by `MadGraph` [25] as inputs; [2] the events are then passed to `Pythia` [26] for parton showering and hadronization, and `Delphes` [27] for detector level object reconstruction, which are also done internally in `CheckMATE`; finally the upper limit is set by statistically evaluation of signal events passed the pre-defined cuts using the CL_S method [28].

This is a typical procedure to test a model against LHC data; the discrimination between signal and background events is performed by several pre-defined signal cuts. On the other hand, it can also be done by Machine Learning (ML) methods. Machine Learning is an algorithm that is able to learn experience from data, which has a great success in many tasks, such as image classification, neural machine translation, etc [29, 30]. A brief introduction is given in Appendix A. Similar to cut-based method, the detector level object reconstruction is required; and we need to extract physical variables like four-momenta from a given event as features. The difference is that, instead of pre-defining cuts, we train a ML classifier to distinguish signals from backgrounds. Since the output of ML classifier is a probability which represents a given event is signal or not, we could set a threshold as "cut" so that any event with a higher probability is a signal. Finally, the upper limit is obtained based on null hypothesis test by assuming a Poisson distribution centered at the events number background. The advantage of ML methods is that we don't need to pre-define different cuts; and it is easy to generalize to other models. However, unlike cut-based method, ML is somehow a "black box", which makes it difficult to explain. But we will try to uncover it a bit in this thesis.

1.4 Structure of the Thesis

This thesis is structured as follows.

In Chapter 2, we give a detailed review on Dark Matter problems. It covers different aspects of evidences that indicates the existence of dark matter. Meanwhile, various experiments searches on dark matter are introduced, including direct detection, indirect detection and collider searches.

In Chapter 3, firstly we discuss the $U(1)$ gauged extension theory of the SM, in particular the anomaly free $U(1)_{L_\mu-L_\tau}$ model. Then, we apply LHC data to constrain this model based on 2-leptons, 3-leptons and 4-leptons final states. The upper bound is derived by using the `CheckMATE` framework.

In Chapter 4, we apply serval Machine Learning algorithms, including Gradient Boosting Decision Tree and Neural Network, to distinguish signals of $U(1)_{L_\mu-L_\tau}$ model from the SM using simulated LHC data. In this way, we find greatly improved sensitivity which exceeds the combination of published LHC and non-LHC results.

[1] Here B is the Baryon number, L is the lepton number and L_i is the i lepton number.

[2] `CheckMATE 2` integrates `MadGraph` into its evaluation process; hence the user only need to pass an `SLHA` file or a `UFO` model file, then events will be generated internally.

Finally, in Chapter 5, we summarize all the work in this thesis.

CHAPTER 2

Dark Matter

In previous chapter, we have briefly mentioned that, the dominant matter in our Universe is not the ordinary matter. Instead, it is the dark matter (DM), which is generally considered to be non-baryonic and contributing to nearly 85% of the total matter. The nature of dark matter is still unknown, but it surely has no electromagnetic interaction, so traditional photometry cannot detect their existence, and it is the reason of being “dark". The only experimental evidences are from gravitational effects, which we will discuss extensively in next section. However, although dark matter is not been directly observed in laboratory, there is no particular reason to prevent them interacting weakly with ordinary matter. This “weak" interaction can be either within the SM or BSM, hence leads to different kinds of dark matter candidates. Among them, a well-studied one is the Weakly Interacting Massive Particle (WIMP). It is well-known for the “WIMP miracle", since it can predict the correct dark matter relic density naturally with masses at weak scale; it is a “miracle" in the sense that there are many well-motivated candidates, for example the lightest supersymmetric particle (LSP) in SUSY.

If dark matter indeed interact weakly with ordinary matter, like nuclei, then they can be detected by experiments of scattering dark matter off nuclei. This is called the direct detection. Since they are weakly interacted, to avoiding heavily noisy background, these experiments are usually built deep underground with low temperature detectors, such as XENON1T [31, 32]. Another kind of DM search is the indirect detection. When DM collide, they can annihilate into ordinary particles; this happens more likely in abundant DM density region, for example the galactic center of the Milky Way Galaxy [33].Then the produced particles can be detected on the surface of Earth by telescopes. Lastly, DM can also be produced in colliders; this is the so-called collider detection. Depending on the actual DM mass, in order to produce DM, the collider energy should be high enough. The LHC has reached a energy level of 13 TeV, which makes it a potential DM factory.

The history of DM is probably longer than we expect, according to Gianfranco Bertone and Dan Hooper’s review [34]. Lord Kelvin was the first one attempted to give an estimation of the amount of DM in the Milky Way. He described the stars in the Milky Way as a gas of particles, bound by gravity. Then, based on the velocities of stars, he obtained an estimation of the mass of the galaxy. But, surprisingly, it was much larger than we observed, so he argued that “Many of our stars, perhaps a great majority of them, may be dark bodies.". Later in 1906, Henri Poincaré explicitly used “dark matter" to describe these invisible matter. However, he disagreed with Lord Kelvin’s conclusion, since there might be unknown obscure stars circulating in the interstellar space. In 1915, Ernst Öpik also concluded that it was unlikely to have large amount of invisible matter. Until the early

1930s, DM was back to public by Jan Oort's work. He studied the motions of stars in the solar neighborhood by Doppler shifts, and he found the calculated velocities would allow them to escape from the gravitational constraint of visible matter. Therefore, he explained it as additional dark matter existed, or the measurements of velocities were simply in error. In 1933, Fritz Zwicky obtained a similar result by studying the Coma cluster. He applied the virial theorem to estimate the cluster's mass, which is the first time to our knowledge. He found that the total mass of the cluster is much larger than the luminous mass, and he concluded that "dark matter is present in much greater amount than luminous matter". However, his explanation of DM still remained in the context of cool and cold stars, solid bodies, and gases.

These are the early attempts to prove the existence and estimate the amount of DM. As we have seen, the conclusions are controversial in this phase. However, with more and more evidences coming out, the existence of DM is getting unambiguous.

2.1 Evidence

2.1.1 Rotation Curves of Spiral Galaxies

It was the rotation curves of spiral galaxies which convinced most of the science community that large amounts of DM are present [34]. The spiral galaxies are described by their spiral arms which extend from the center into outside flat rotating disk. At larger distance from the dense center, if we only consider gravitational force, the rotation velocity should be the following form,

$$\frac{GM}{r^2} = \frac{mv^2}{r} \implies v = \left(\frac{GM}{r}\right)^{1/2} . \tag{2.1}$$

where G is the gravitational constant, M is the galactic mass within radius r. This means that if the luminous matter are all the matter existed in the galaxy, the velocity should be decreased with radius as $v \propto r^{1/2}$. However, it is certainly not the case observed by Vera Rubin and his collaborators in 1970s [35].

By studying the rotation curve of the M31 galaxy, Vera Rubin [35] found the velocity distribution was extremely deviate from Newton's laws. Instead of being convex, the rotation curve has a "flat" shape, which means that the velocities of distant stars continue to increasing until reaching constant. Following observations of other galaxies also conformed Rubin's results, such as the rotation curve of NGC 3198 [36], as shown in Fig. 2.1. The luminous matter is concentrated near the galactic center, whereas there is little matter on the outside flat disk. If the collected data is corrected, and Newton's laws are respected, additional invisible matter must be added while increasing the radius. So that the actual mass enclosed in the radius r is like $M(r) \propto r$, leading to a flat curve in Fig. 2.1. This implies that the majority of matter in spiral galaxies is DM.

2.1.2 Gravitational Lensing

The four dimensional space-time which we live in is described by Einstein's general theory of relativity. Any massive object would bend the space, and alter the motions of other objects, even the light. This effect is called the gravitational lensing. To see such effects, the target object should be massive enough to bend the path of light; usually they are cluster of galaxies. Moreover, to amplify the bending

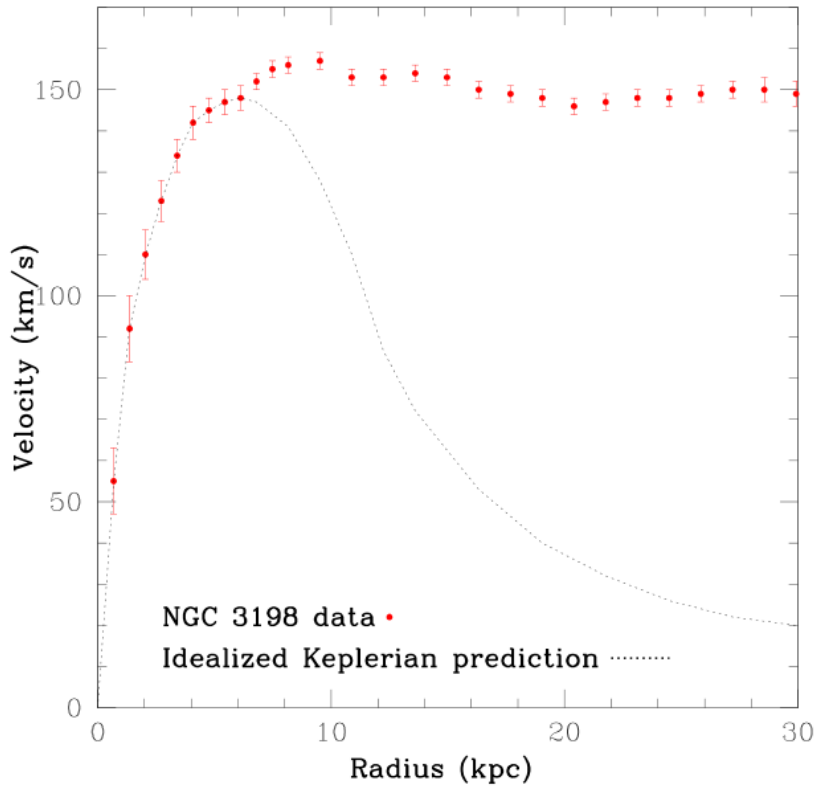

Figure 2.1: Measured rotation curve of NGC 3198 [22, 36]. The red-dot is the measured data, while the black-dashed line is the predicted value.

arcs, cosmologists often look for the relatively close target which lies almost exactly between the observer (the Earth) and the distant, bright galaxies. By studying the distorting images of distant galaxies, we can estimate how massive the target cluster is to produce such a bending. However, it turns out that the calculated mass is much larger than the luminous mass we observed. This indicates large amount of DM exist in cluster of galaxies. One of such results is shown in Fig. 2.2, two images of the massive galaxy cluster Cl 0024+17 (ZwCl 0024+1652) from Hubble's view[1]. The left one shows the visible matter, the odd-looking blue arcs appearing among the yellowish galaxies are the magnified and distorted images of distant galaxies. The blue shading region in the right figure is the calculated DM distribution which can explain the observed gravitational lensing.

2.1.3 Bullet Cluster

Evidence from the Bullet cluster is considered as the "smoking-gun" for existence of DM. The Bullet cluster (1E 0657-56) consists of two main colliding clusters of galaxies, the centers are separated by about one megaparsec, or 3.26 million light years [22]. The major components are similar in these

[1] from NASA's official website: https://www.nasa.gov/content/discoveries-highlights-shining-a-light-on-dark-matter.

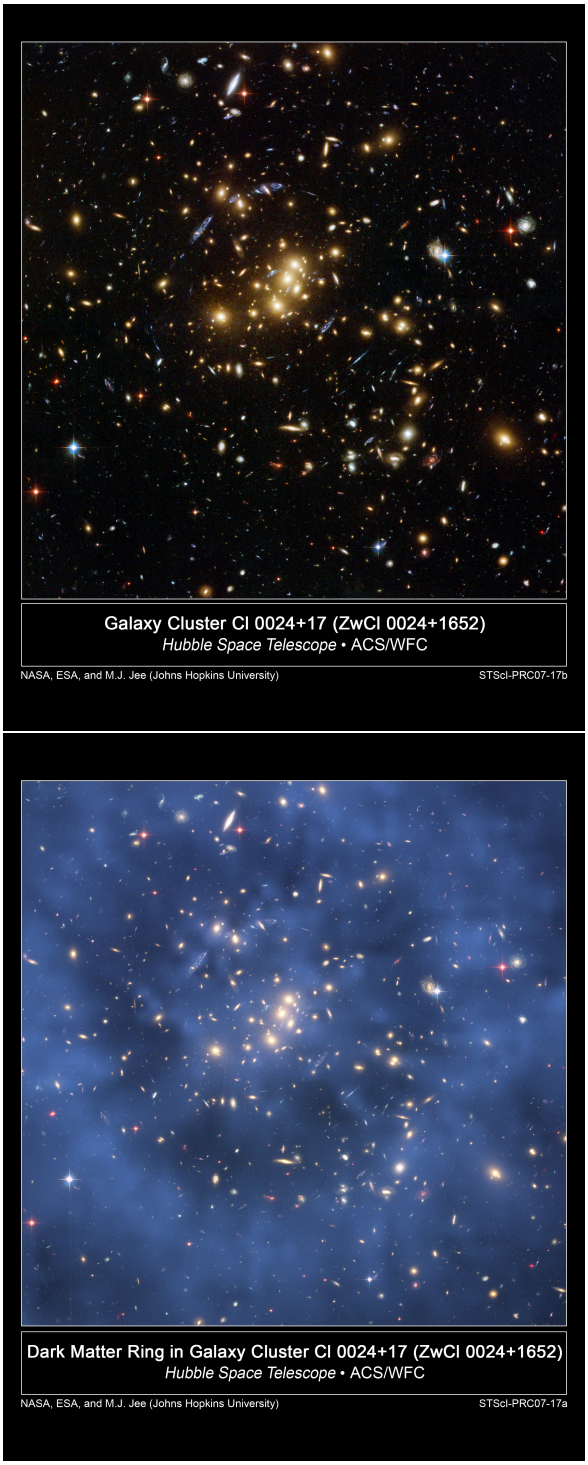

Figure 2.2: Two views from Hubble of the massive galaxy cluster Cl 0024+17 (ZwCl 0024+1652) from NASA's website. The upper frame is the view in visible-light; the blue shading in the lower frame indicates the location of invisible dark matter.

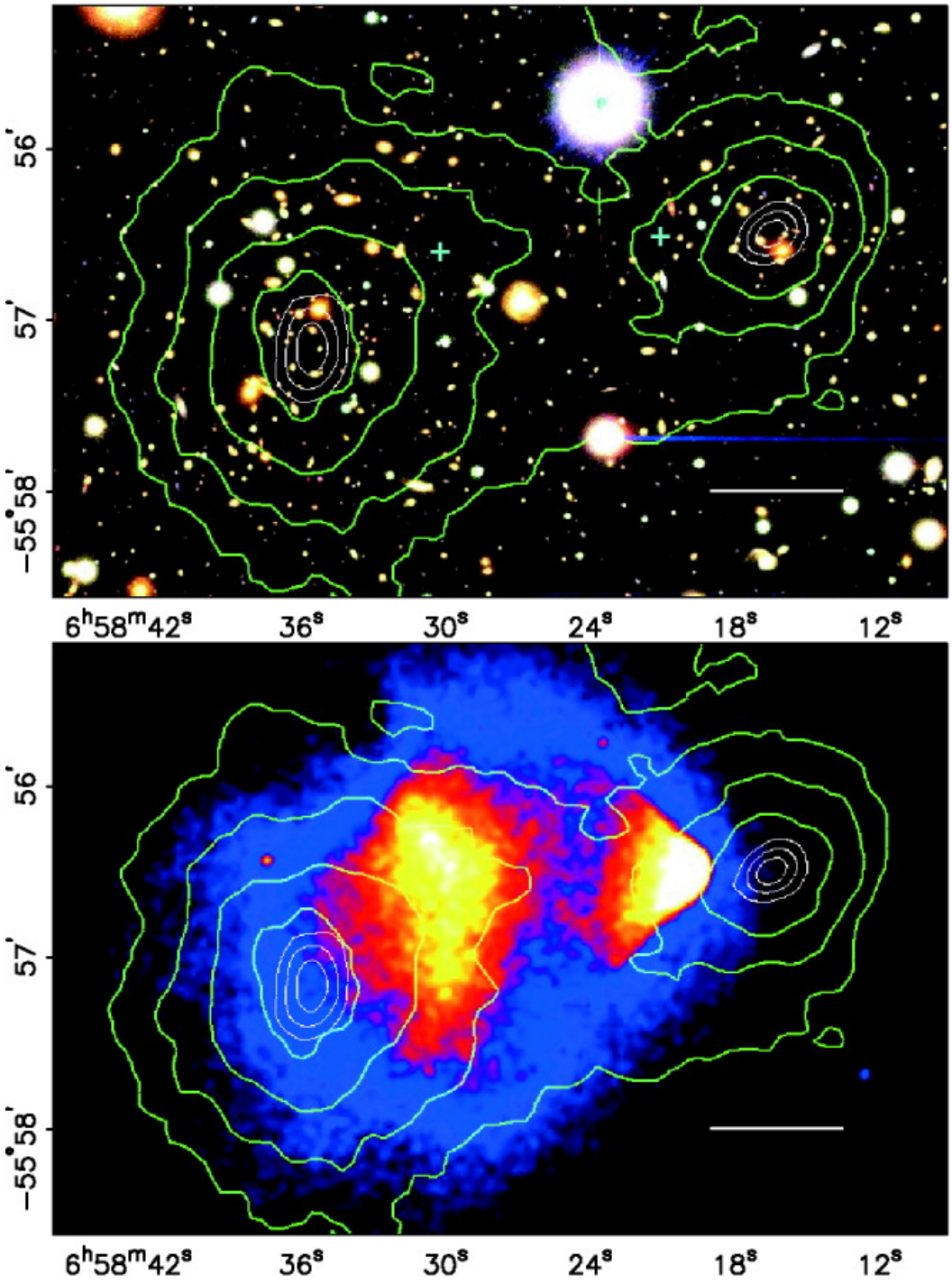

Figure 2.3: Weak gravitational lensing contours (green) and hot gases (red) in Bullet cluster [37].

two subclusters, which are galaxies, gas and possibly DM. During collision, their mass distributions behave differently with or without DM. The two galaxy centers are not affected by the collision, since they are separated far enough. However, the hot gas will be compressed and heated during collision, which results in a large amount of X-ray radiations. They are shown in Fig. 2.3, where the red region indicates the hot gas. Moreover, since the hot gas are interacting with each other, they will be slowed down by collisions. If DM do not exist, and all matter are visible and interacting with electromagnetic fields, then the gravitational mass distribution should concentrate in the X-ray region. However, it is found that the major mass distributed around the collisionless galaxies. This clearly points to the existence of DM, and rejects assumptions of modified Newton Dynamics.

2.1.4 Cosmic Microwave Background (CMB)

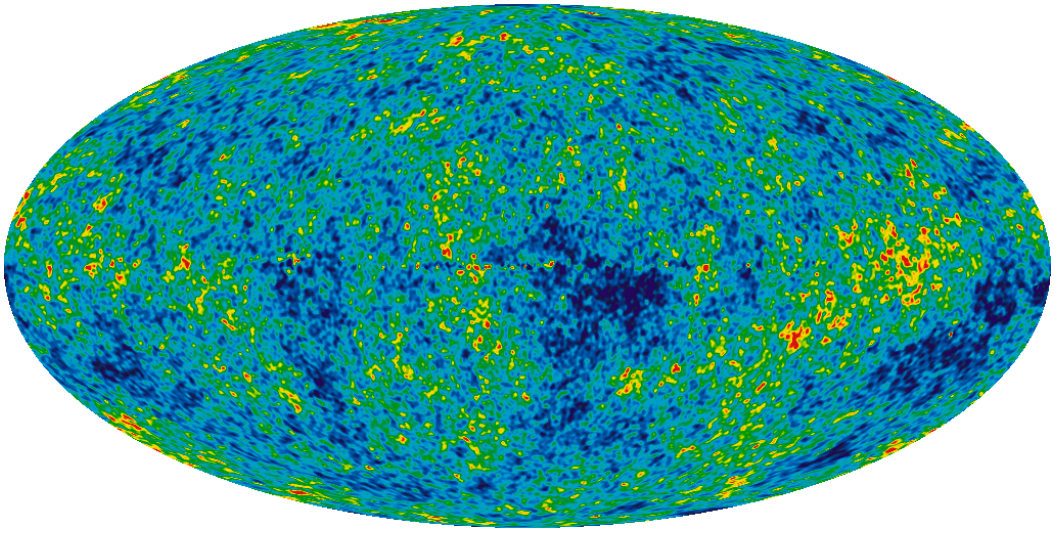

Figure 2.4: WMAP's map of the temperature of the CMB, with hot regions in red and cooler regions in blue [38].

According to the standard cosmology, a few seconds after the Big Bang, the Universe was filled with photons and charged particles, such as electrons and positrons. It was like a hot sea, and all these particles inside it were considered to be in thermal equilibrium. As time went on, the universe expanded and cooled for about about $380,000$ years, then it reached the epoch of recombination. At this time period, the particles fell out of equilibrium, protons and neutrons formed nuclei, and free electrons were absorbed to form neutral atoms. Hence, the universe became transparent, and the photons were left to be free, which are now known as the cosmic background radiation. The Wilkinson Microwave Anisotropy Probe (WMAP) gave a precise measurement of CMB, they found that the temperature of the microwave radiation is not uniformly distributed, the difference is small, about 30μK, as shown in Fig. 2.4. One popular explanation is from the Lambda-CDM (Lambda cold dark matter) model. Since no interactions with electromagnetic field, the DM could be formed long before the recombination epoch. Hence, while the neutral atoms formed, they were experienced a gravitational pull by DM, and same happened with photons. Then, through the CMB anisotropies, we were able to estimate the number density of DM via fitting the temperature power spectra. The latest

result is from the Planck 2018 data [39],

$$\Omega_c h^2 = 0.120 \pm 0.001, \quad \Omega_b h^2 = 0.0224 \pm 0.0001\,. \tag{2.2}$$

where $\Omega_c h^2$ is the DM density, and $\Omega_b h^2$ is the baryon density. From this data, we can get the DM contributed 84.2% to the total matter.

2.2 Direct Detection

If DM is abundant in the Universe, and they are weakly interacted with ordinary matter, then they can be detected by the scattering process with nucleus, which is the direct detection [40]. Among all of the DM candidates, the Weakly Interacting Massive Particles (WIMPs) are one of the most well-studied. Hence, we will focus on the WIMPs detections in this and following sections.

2.2.1 The Basics

The expected signals of direct detection depend on the distribution and density of DM in the Milky Way. The typical local DM density ρ_0 is 0.3 GeV/cm^3, but with a large uncertainty across different measurements. In the Standard Halo Model, the velocity distribution of DM follows the solution of the Boltzmann equation for collisionless particles [40],

$$f(v) = N \exp\left(-\frac{3|v|^2}{2\sigma^2}\right)\,. \tag{2.3}$$

where N is the normalization constant, σ is the velocity dispersion. The average DM velocity is around 220 km/s. It is also important to notice that, the velocity of DM could go to infinite. But if it's large enough, such as larger than the escape velocity v_{esc}, then DM are not bounded by the gravitational effect. So there is a maximum speed for DM in direct detection, in a range of 498 − 608 km/s [40]. Then, the expected event rate of WIMPs scattering is estimated as [40],

$$\frac{dR}{dE_{nr}} = \frac{\rho_0 M}{m_N m_\chi} \int_{v_{min}}^{v_{esc}} \frac{d\sigma}{dE_{nr}} dv\,. \tag{2.4}$$

where E_{nr} is the nuclear recoil energy, m_N and m_χ is the mass of target nucleus and WIMP. M is the target mass of the detector, and v_{min} is the minimum velocity of DM required to make a E_{nr} nuclear recoil. The scattering cross section could be further divided into Spin-Independent (SI) and Spin-Dependent (SD) cross section [40],

$$\frac{d\sigma}{dE_{nr}} = \frac{m_N}{2v^2\mu^2}(\sigma_{SI} F_{SI}^2(E_{nr}) + \sigma_{SD} F_{SD}^2(E_{nr}))\,. \tag{2.5}$$

μ is the reduced mass, F is the form factor. The SI cross section is proportional to the nuclear atomic number, whereas the SD cross section is proportional to the total spin. Hence, SI scattering usually has a large event rates, but SD scattering are more prone to detect vector interactions.

2.2.2 Experimental Results

Although numerous experiments of direct detection have been looking for DM for decades, including but not limited to LUX [41], PandaX-II [42], DarkSide [43], SuperCDMS [44], XENON1T [45], there is no strong level of significance indicating the existence of DM. So the coupling between DM and baryonic matter must be weak enough to evade current search. The current upper bound for WIMP SI direct detection is shown in Fig. 2.5. The dashed origin line is the “neutrino floor", it is a limit from the neutrino-nucleus scattering background. The strongest bound with WIMP mass above 5 GeV is from XENON1T, while for WIMP mass below 5 GeV, it is DarkSide-50. In the case of SD direct detection, since the cross section is dependent on the overall nucleus' spin, neutrons and protons have different contributions to the total cross section. The corresponding bound is shown in Fig. 2.6. The dominant constraint is again from XENON1T, due to its abundant unpaired neutrons.

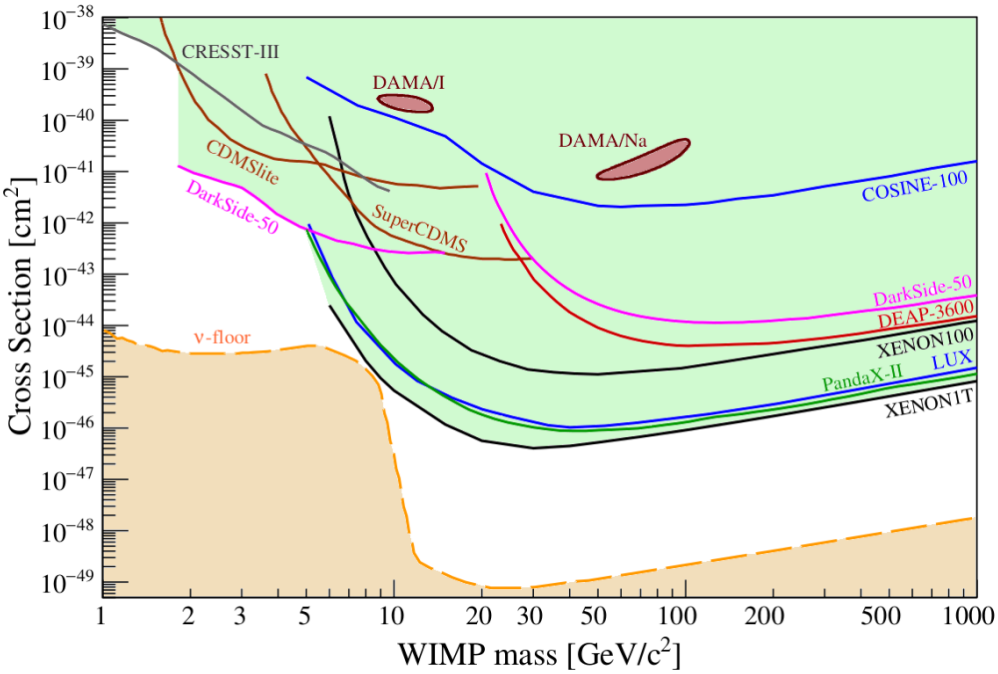

Figure 2.5: The current upper bound for SI WIMP-nucleon cross section, figure from [40].

2.3 Indirect Detection

Indirect detection [46] is based on searching for excess particles produced by DM annihilations or decays in the Universe. These particles could be gamma rays, cosmic rays, and neutrinos. Among them, the flux of neutral products is [47],

$$\Phi(E,\phi) = \frac{\Gamma}{4\pi m_\chi^a}\frac{dN}{dE}\int \rho[r,(l,\phi)]^a dl\,. \tag{2.6}$$

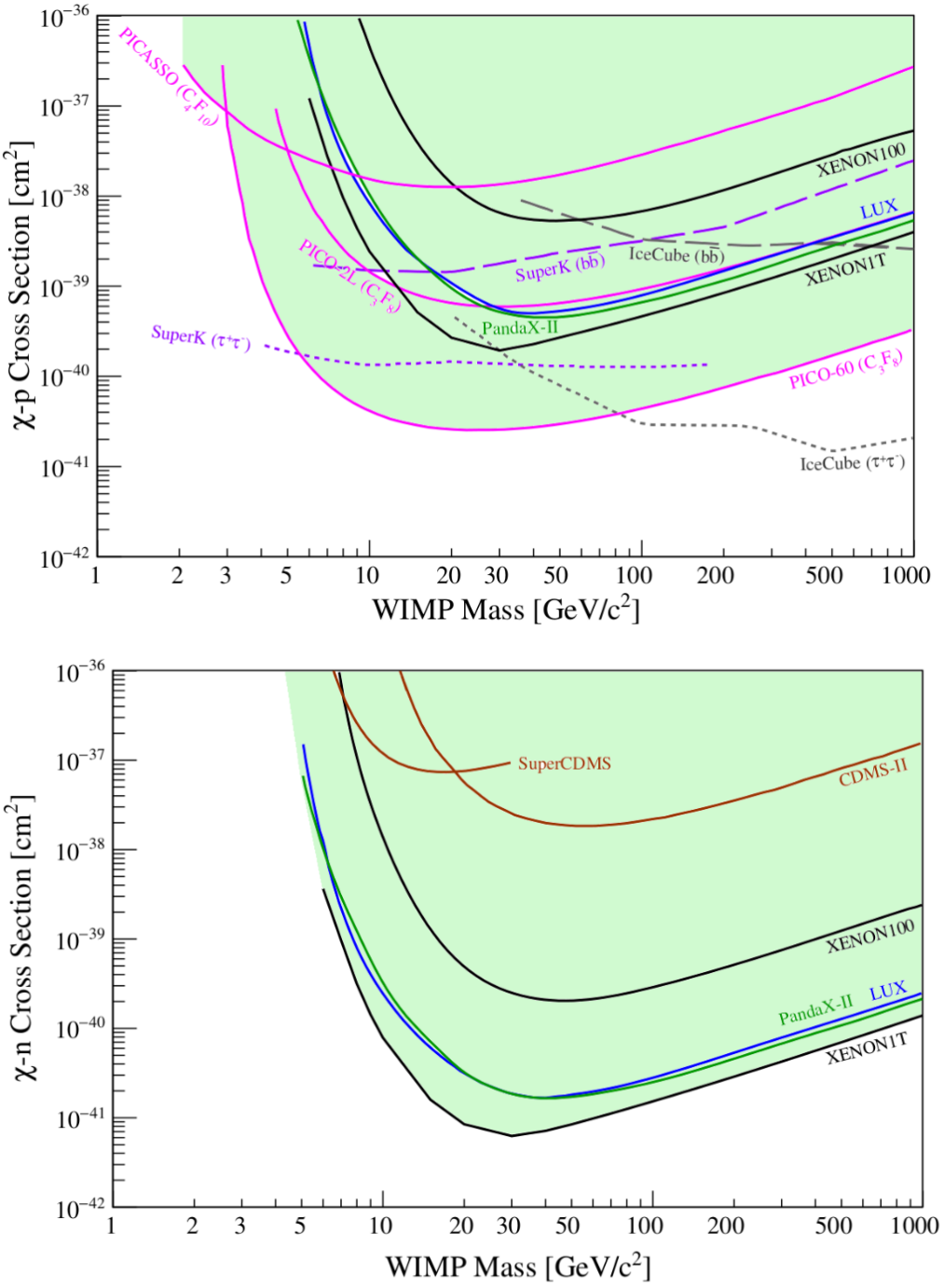

Figure 2.6: Current status of SD searches. (Top) WIMP-proton interactions. (Bottom) WIMP-neutron interactions [40].

where Γ is the DM interaction rate, and ρ is the DM density. For annihilations, two DM collide and $a = 2$, Γ is proportional to the corresponding annihilating cross section, which is $\Gamma = \langle\sigma v\rangle/2$. For DM decays, $a = 1$ and Γ is the inverse of DM lifetime τ_t, $\Gamma = 1/\tau_t$. Unlike direct detection, indirect detection is not limited to the DM mass or couplings, it can probe much broad range of energies and decay lengths of DM. However, it also suffers from many unknown backgrounds. To optimize the signals, physicists usually looking for scenarios with enhanced channels like Sommerfeld enhancement, and places with high DM density.

Among of the annihilating or decaying products, gamma rays can be produced by DM annihilation directly, or emitting by DM final products. Moreover, due to their neutral properties, they can travel through galaxies without being affected by magnetic fields. Hence, in general, gamma rays could be detected with strong signals and set stringent bound on DM annihilation cross section, as shown in Fig. 2.7. A promising finding is the $2 - 3$ GeV excess detected by Fermi Gamma-Ray Space Telescope [48], it could be fitted well by WIMP with mass between $50 - 100$ GeV, and annihilation cross section $\langle\sigma v\rangle \sim 10^{-26}$ cm^3/s. However, it could also be produced by indistinguishable pulsars backgrounds, and the explanation still remains inclusive.

2.4 Collider Searches

Unlike direction and indirect detection we discussed above, collider searches [51] don't depend on the galactic DM. Instead, they are looking for signals of DM produced by SM particles in colliders, such as Large Electron Positron Collider (LEP) and LHC. In this thesis, we will focus on LHC, specifically CMS and ATLAS. Due to their weak couplings to SM particles, DM leaves no trace in colliders and is accounted as missing energy. By studying the excess missing transverse momentum, collider searches are able to study a wide range of interactions between DM and SM particles.

In the early days of collider search, searching for specific new physics models like Supersymmetry (SUSY) is one of its main topics. SUSY with R-parity can avoid lepton number and baryon number violation, and it can also lead to a stable WIMP candidate, named the lightest SUSY particle (LSP). The relevant signals for searching LSP are multiple jets, accompanied by leptons and missing energy. However, due to the complexity and large number parameter space of SUSY, no evidences have been found so far.

In recent years, collider experiments have changed the directions to searching for simplified and generalized models, mediated by vector bosons or scalar particles. Depending on the nature of mediators, the relevant signals could be multiple leptons or jets with missing energy. For example, in the model of $U(1)_{L_\mu - L_\tau}$, which is the main topic in this thesis, the new gauge boson Z' serves as the mediator. It has non-vanishing couplings to both SM particles (like muons and tauons) and invisible DM. The couplings and masses are free parameters in this model, and are constrained by excess events of leptons and missing energy. We will discuss its phenomenology in more detail in the following chapters.

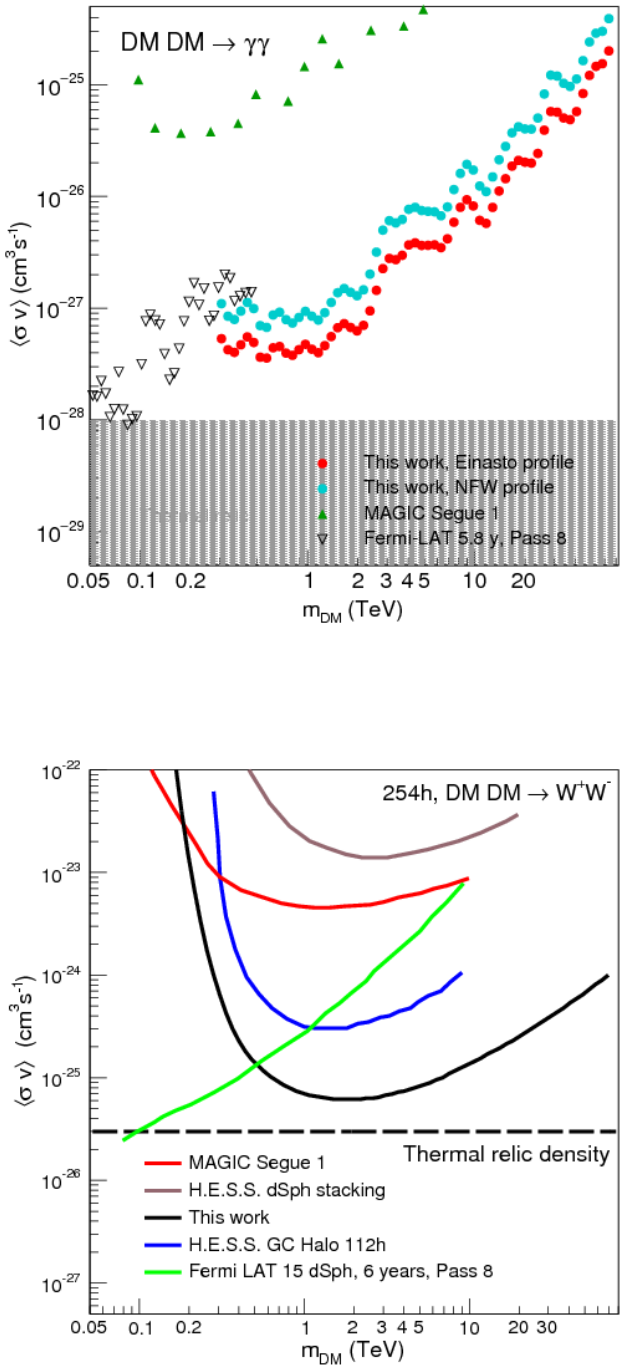

Figure 2.7: (Top) Galactic center gamma-ray line limits from H.E.S.S. (Bottom) Annihilation to WW limits from H.E.S.S. Plot is taken from [47, 49, 50].

CHAPTER 3

Constraints on $U(1)_{L_\mu - L_\tau}$ from LHC Data

3.1 U(1) Gauged Extensions

We have discussed the limits of SM in previous chapters, such as neutrino mixing and lack of viable dark matter candidate. There are many possible extensions of SM to solve partial of these problems. Among of them, U(1) gauged extension is possibly the simplest one. It extends the SM gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$ to $SU(3)_C \times SU(2)_L \times U(1)_Y \times U(1)_X$, where $U(1)_X$ represents an additional gauge symmetry. Since the Baryon number (B) and the lepton number (L) are conserved in SM, a natural guess of X would be B or L. However, a naive local gauge of $U(1)_B$ or $U(1)_L$ leads to an anomaly theory. In order to get an anomaly free theory, we need to consider some combinations between B and L. There are two possible cases:

- We consider the difference between B and L, $U(1)_{B-L}$, and introduce extra chiral fermions such as right handed neutrinos to cancel gauge anomaly.

- We consider the difference between lepton number of different generations, which are $L_e - L_\mu$, $L_e - L_\tau$ and $L_\mu - L_\tau$. L_e, L_μ and L_τ are the lepton number of e, μ and τ generation. We can generalize it into $L_\alpha - L_\beta$, where $\alpha(\beta) = e, \mu, \tau$. Unlike $B - L$, we don't need extra fermions to cancel anomaly in this case.

To break the extra $U(1)_X$ gauge, we need further introduce a Higgs-like complex scalar ϕ_H with non-trivial $U(1)_X$ charge. During the process of spontaneous symmetry breaking, the new gauge boson Z' could mix with SM electroweak (EW) bosons, depending on whether it couples to SM gauge bosons or not. But for simplicity, we only consider the minimal case with no coupling to EW sector. After the spontaneous symmetry breaking, Z' will acquire a mass $m_{Z'}$ proportional to v_H the vacuum expectation value of ϕ_H. Besides, we introduce three righted neutrinos N_e, N_μ, N_τ to generate neutrino mixing in both cases, although it's not a necessary part in case of $L_\alpha - L_\beta$. We can also enlarge the model with a dark sector by adding an extra complex scalar ϕ_{DM}, which is charged under the $U(1)_X$ gauge. Also we set this scalar field ϕ_{DM} odd under a $\mathbb{Z}_2$ symmetry to make it a stable and a viable dark matter candidate. With these additional fields and proper assignment of new $U(1)_X$ charges, we can develop a simple extension of SM, which will be discussed in following sections.

3.1.1 $U(1)_{B-L}$

With these additional fields mentioned above, the Lagrangian of $U(1)_{B-L}$ model is given by,

$$\mathcal{L} = \mathcal{L}_{SM} + \mathcal{L}_N + \mathcal{L}_{DM} + (D_\mu \phi_H)^* (D^\mu \phi_H) - V(\phi_h, \phi_H) - \frac{1}{4} F'^{\mu\nu} F'_{\mu\nu} \,. \tag{3.1}$$

where $\mathcal{L}_{SM}$ is the Lagrangian of SM, and ϕ_h is the SM Higgs doublet to spontaneously break electroweak symmetry. $\mathcal{L}_{DM}$ is the Lagrangian of the dark sector, which includes interaction terms involving ϕ_{DM}.

$$\begin{aligned} \mathcal{L}_{DM} = (D_\mu \phi_{DM})^* (D^\mu \phi_{DM}) - \mu_{DM}^2 \phi_{DM}^* \phi_{DM} - \lambda_{DM} (\phi_{DM}^* \phi_{DM})^2 \\ - \lambda_{Dh} (\phi_{DM}^* \phi_{DM})(\phi_h^\dagger \phi_h) - \lambda_{DH} (\phi_{DM}^* \phi_{DM})(\phi_H^* \phi_H) \,. \end{aligned} \tag{3.2}$$

The scalar potential $V(\phi_h, \phi_H)$ containing interactions between ϕ_h and ϕ_H is,

$$V(\phi_h, \phi_H) = \mu_H^2 \phi_H^* \phi_H + \lambda_H (\phi_H^* \phi_H)^2 + \lambda_{hH} (\phi_h^\dagger \phi_h)(\phi_H^* \phi_H) \,. \tag{3.3}$$

Note that the potential term $V(\phi_h, \phi_H)$ should also contain the self interaction of ϕ_h, which is $\mu_h^2 \phi_h^\dagger \phi_h + \lambda_h (\phi_h^\dagger \phi_h)^2$, but they are already in SM Lagrangian $\mathcal{L}_{SM}$. The potential $V(\phi_h, \phi_H)$ must be bounded from below, which implies that the following conditions must be hold [52],

$$4\lambda_h \lambda_H - \lambda_{hH}^2 > 0 \,, \quad \lambda_h, \lambda_H > 0 \,. \tag{3.4}$$

Then the $SU(2)_L \times U(1)_Y \times U(1)_{B-L}$ gauge symmetry breaks spontaneously while ϕ_h and ϕ_H acquire VEVs, and the corresponding gauge bosons become massive. In unitary gauge, ϕ_h and ϕ_H can be expressed as,

$$\phi_h = \begin{pmatrix} 0 \\ \frac{v_h + h}{\sqrt{2}} \end{pmatrix}, \quad \phi_H = \left(\frac{v_H + H}{\sqrt{2}}\right) . \tag{3.5}$$

where v_h and v_H are corresponding VEVs of ϕ_h and ϕ_H, h and H are remaining scalar fields. Since the presence of interaction coupling λ_{hH}, these two fields will mix up with each other. Then the physical states h_1 and h_2 in mass eigenstates are [52],

$$\begin{aligned} h_1 &= h \cos\alpha - H \sin\alpha \,, \\ h_2 &= h \sin\alpha + H \cos\alpha \,, \\ \tan\alpha &= \frac{\lambda_{hH} v_h v_H}{\lambda_h v_h^2 - \lambda_H v_H^2} \,. \end{aligned} \tag{3.6}$$

Their masses m_{h_1} and m_{h_2} can be written as [52],

$$\begin{aligned} m_{h_1} &= \lambda_h v_h^2 + \lambda_H v_H^2 - \sqrt{(\lambda_h v_h^2 - \lambda_H v_H^2)^2 + (\lambda_{hH} v_h v_H)^2} \,, \\ m_{h_2} &= \lambda_h v_h^2 + \lambda_H v_H^2 + \sqrt{(\lambda_h v_h^2 - \lambda_H v_H^2)^2 + (\lambda_{hH} v_h v_H)^2} \,. \end{aligned} \tag{3.7}$$

It is obvious that $m_{h_1} < m_{h_2}$, and if we consider h_1 as the SM Higgs boson, then its mass $m_{h_1} = 125$ GeV and VEV $v_{h1} = 246$ GeV. After spontaneously symmetry breaking, DM candidate ϕ_{DM} acquires its

mass from both fields ϕ_h and ϕ_H,

$$m_{DM} = \mu_{DM}^2 + \frac{\lambda_{Dh} v_h^2}{2} + \frac{\lambda_{DH} v_H^2}{2} . \tag{3.8}$$

The Lagrangian $\mathcal{L}_N$ which contains the kinetic terms and Yukawa terms involving the extra right handed neutrinos N_i can be expressed as [52]

$$\mathcal{L}_N = \sum_{i=e,\mu,\tau} (\bar{N}_i \gamma^\mu D_\mu N_i - y_i^\nu \bar{L}_i \tilde{\phi}_h N_i - y_i^N \bar{N^c} N \phi_H) + \text{h.c.} \tag{3.9}$$

where $\tilde{\phi}_h = i\sigma_2 \phi_h^*$, y^ν and y^N are dimensionless Yukawa couplings. The last term $\bar{N^c} N \phi_\chi$ corresponds to the Majorana mass term. Furthermore, the covariant derivative D_μ which appears in $\mathcal{L}_{SM}$ and $\mathcal{L}_N$ is defined as

$$D_\mu \equiv \partial_\mu + i g_3 T_3^a G_\mu^a + i g_2 T_2^a W_\mu^a + i g_1 Y B_\mu + i g' Q_{B-L} B'_\mu . \tag{3.10}$$

Here we consider no mixing between the new gauge boson Z' and SM bosons. All the fermions are coupled with Z', and the couplings are g'. However quarks and leptons have different charge under $U(1)_{B-L}$, in particular, quarks have charge of $1/3$ and leptons of -1. Therefore the extra scalar field ϕ_H is charged by +2 to make the Lagrangian gauge invariant.

The phenomenology consequences of $U(1)_{B-L}$ model cover different aspects. Firstly, with the introduction of extra neutrinos, neutrinos masses are generated by seesaw mechanism, so it is constrained by neutrinos mixing experiments. Secondly, the extra scalar field ϕ_H leads to a two-Higgs sector, and extra higgs has been searched extensively in different experiments [53–55]. Lastly but not least, the new gauge boson Z' couples both to leptons and quarks, which would suffer great constraints from electron-positron collider (LEP) and hadron collider (LHC). A detailed analysis of $B - L$ constraints can be found in [56].

3.1.2 $U(1)_{L_\alpha - L_\beta}$

With the additional fields of three right handed neutrinos and scalar dark matter fields, the Lagrangian of $U(1)_{L_\alpha - L_\beta}$ is similar to $U(1)_{B-L}$ in Eq. (3.1). The difference appears in the charges of covariant derivative term,

$$D_\mu \equiv \partial_\mu + i g_3 T_3^a G_\mu^a + i g_2 T_2^a W_\mu^a + i g_1 Y B_\mu + i g' Q_{L_\alpha - L_\beta} B'_\mu \tag{3.11}$$

Only partial leptons couple to the new gauge boson Z', and different generations have different charges. For example, in the $L_\mu - L_\tau$ case, only the second (μ) and the third generation (τ) are charged under Z', but not the first generation (e). μ has a charge of +1 and τ of -1. This will alter the Lagrangian of

fermions in Eq. (3.9) a bit to make sure the conservation of gauge invariance, which is given by [57].

$$\begin{aligned}\mathcal{L}_N = &\sum_{i=e,\mu,\tau} \frac{i}{2}(\bar{N}_i \gamma^\mu D_\mu N_i) - \frac{1}{2} M_{ee} \bar{N}_e^c N_e - \frac{1}{2} M_{\mu\tau}(\bar{N}_\mu^c N_\tau + \bar{N}_\tau^c N_\mu) \\ &- \frac{1}{2} h_{e\mu}(\bar{N}_e^c N_\mu + \bar{N}_\mu^c N_e)\phi_H^\dagger - \frac{1}{2} h_{e\tau}(\bar{N}_e^c N_\tau + \bar{N}_\tau^c N_e)\phi_H \\ &- \sum_{i=e,\mu,\tau} y_i^\nu \bar{L}_i \tilde{\phi}_h N_i + \text{h.c.}\,.\end{aligned} \tag{3.12}$$

After spontaneously symmetry breaking, neutrino masses are diagonalized through seesaw mechanism, right handed neutrinos are heavy and left handed ones are light. The masses of two Higgs like scalars h_1, h_2 and the dark matter candidate ϕ_{DM} also take the form as in Eq. (3.7) and Eq. (3.8).

The phenomenology of $L_\alpha - L_\beta$ is also similar to $B - L$, which includes neutrino mixing, extra higgs sector and dark sector, as well as the Z' decay channel. However, for models involving extra couplings to electrons, like $B - L$, $L_e - L_\mu$ and $L_e - L_\tau$, LEP sets strong constraints on the new gauge boson [58, 59].

$$\begin{aligned}\frac{m_{Z'}}{g'} &> 6.9 \text{ TeV} \qquad B - L\,, \\ \frac{m_{Z'}}{g'} &> 5.25 \text{ TeV} \qquad L_e - L_\mu, L_e - L_\tau\,.\end{aligned} \tag{3.13}$$

The remaining gauge group is $L_\mu - L_\tau$, which is also the main topic of this thesis. It escapes from constraints at LEP, but still suffers from the neutrino mass bounds and searches of new gauge boson at LHC. The detailed discussion on the constraints of $B - L$ and $L_\alpha - L_\beta$ models can be found in [60, 61].

3.2 Simplified $L_\mu - L_\tau$ Model

We have discussed that, for models like $B - L$, $L_e - L_\mu$ and $L_e - L_\tau$, the new gauge boson couples to charged leptons, which makes them tightly constrained by di-lepton resonance searches at lepton and hadron colliders. In contrast, $L_\mu - L_\tau$ model does not add any new tree-level interaction with electrons. The new gauge boson can only be produced by high-order processes, for example by emission from a tauon. Hence it is much less constrained by current experiments. Besides, with the additional scalar dark matter field ϕ_{DM}, it can provide a viable DM candidate which can easily escape current direct search constraints [62, 63]. The model can also provide a loop correction to muon's magnetic dipole moment to explain the $g_\mu - 2$ anomaly; however, the new gauge bosons are constrained by trident production experiments [62, 63], which exclude the masses $m_{Z'} > 0.5$ GeV.

The formulism of $L_\mu - L_\tau$ has been discussed in last section, but we make some simplifications here. The Lagrangian is in Eq. (3.1), we specify the covariant derivative in Eq. (3.10) as $D_\mu \equiv \partial_\mu - i g_{\mu\tau} q B'$, where $g_{\mu\tau}$ is gauge coupling and q is the $L_\mu - L_\tau$ charge. The scalar fields ϕ_{DM} and ϕ_H are charged under the new gauge group with charge q_{DM} and q_H. In our model, q_{DM} is a free parameter but q_H is fixed at 1. The extra Higgs sector could have a significant contribution in DM phenomenology [62, 63], but this only happens when the mixing angle α is relatively large. We neglect this possible contribution and set $\alpha = 0$. Therefore the remaining free parameters in our model are $m_{Z'}$, $g_{\mu\tau}$ and q_{DM}.

The $L_\mu - L_\tau$ model predicts a new gauge boson Z', which can be produced by high-order processes with emission of muons (μ or ν_μ) or tauons (τ or ν_τ). These di-muon and di-tau events are abundant at LHC as compared to LEP, we therefore focus on LHC data in this chapter. Most of the existing constraints in the searches for Z' at LHC comes from processes of Z' decaying into four charged leptons [64]; for example, CMS [65] published an analyses which searches for Z' gauge boson using $Z' \to 4\mu$ decays. It is sensitive to relatively light mass of $m_{Z'}$, $m_{Z'} < m_Z$. Besides, $L_\mu - L_\tau$ has already been discussed in [66–68], but they did not take use of LHC data.

In this chapter, we will discuss the LHC constraints on $L_\mu - L_\tau$ model. The corresponding signal final state contains two, three or four charged leptons l, here l is a muon or a hadronically decaying tauon. In the case of two and three leptons, we also need to include the missing transverse momentum $\not{E}_T$, especially when Z' decays into two DM particles. In addition to Z' decays, the process of $\tau \to \mu$ also contributes to final muons, but the resulting muons are obviously softer than parent tauons. In principle, we may also need to consider final state with electrons since the decays of $\tau \to e$. However, since the branching ratio is small (about 18%), its cross section will not be comparable to final states with muons and tauons. Besides, decays of $\tau \to \mu$ will further amplify the cross section of final states with muons, which makes electron decays more negligible. At leading order, the corresponding partial width of Z' are given by,

$$\Gamma(Z' \to l^+ l^-) = \frac{g_{\mu\tau}^2 m_{Z'}}{12\pi}\sqrt{1-4z_l}(1+2z_l) \quad \text{for } l = \mu, \tau\,. \tag{3.14}$$

$$\Gamma(Z' \to \phi_{DM}\phi_{DM}^\dagger) = \frac{q_{DM}^2 g_{\mu\tau}^2 m_{Z'}}{48\pi}(1-4z_{DM})^{3/2} \quad \text{for } l = \mu, \tau\,. \tag{3.15}$$

where $z_X \equiv \frac{m_X^2}{m_{Z'}^2}$. In particular, for the case of decaying into neutrinos, the corresponding partial width is half of which in Eq. (3.14). The reason is that right-handed neutrinos are too heavy to contribute, and only left-handed left in final states. Moreover, in order to make sure that our model is a perturbative theory, we only consider when the total Z' width is smaller than $m_{Z'}$. This means the following condition must be satisfied,

$$q_{DM}^2(1-4z_{DM})^{3/2} + 4\sum_{l=\mu,\tau}\sqrt{1-4z_l}(1+2z_l) + 4 < 48\pi/g_{\mu\tau}^2\,. \tag{3.16}$$

which is always satisfied for $g_{\mu\tau} \leq 3$ and $q_{DM} \leq 2$.

The Feynman diagrams of signal processes that we considered are shown in Fig. 3.1. These are Drell-Yan processes mediated by neutral or charged gauge boson. At the leading order, Z' is produced by emission off the lepton.

In the first diagram (left one), the primary Drell-Yan process mediated by Z boson or photon produces a lepton pair, one of which emitted a Z' boson. If Z' decays into another lepton pair, then the final states contains four charged leptons, which is $pp \to 2l2l'$ ($l, l' \in \{\mu, \tau\}$). These two lepton pairs could be the same flavor or not. On the other hand, if Z' decays invisibly into neutrinos or DM particles, it leads a final state containing an opposite-sign same-flavor(OSSF) lepton pairs and missing energy $\not{E}_T$.

In the second diagram (central one), the primary Drell-Yan process is mediated by a Z boson, and produces a neutrino pair. If Z' decays into charged leptons, then again we get a final state with OSSF

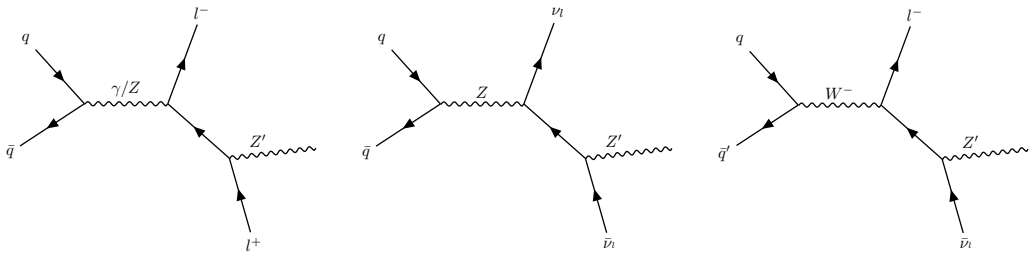

Figure 3.1: Examples of Feynman diagrams for $pp \to Z' l^+ l^-$ (left), $pp \to Z' \nu_l \bar{\nu}_l$ (center) and $pp \to Z' l \nu_l$ (right). In the left diagram, Z' can decay visibly (into leptons) or invisibly (into neutrinos or DM), both of them are considered as signal processes. But for the central and right diagrams, we only consider the case that Z' decays visibly.

leptons and $\not{E}_T$. But, if Z' decays invisibly, it leads to a undetectable final state, which only contains missing energy.

In the last diagram (right one), the primary Drell-Yan reaction produces a $l^- \bar{\nu}$ pair or its charge conjugate through a W boson. Similar to first two cases, the emitted Z' boson can decay either visibly or invisibly. If Z' decays into charged leptons, this leads to a 3-leptons final state $l^\pm l'^+ l'^- \not{E}_T$, where l and l' may again be the same flavor or not. If Z' decays invisibly, the final state is a single charged lepton plus missing energy. This can be considered as a higher-order correction to SM charged Drell-Yan process. We ignore this signal process since it is surely less sensitive than the $3l$ case. Besides, there are no searches for one lepton final state in CMS and ATLAS analyses.

Therefore, the final states we interested are involving either muons or tauons or both and missing energy. Experimentally, muons and tauons are different. However, since the decay of $\tau \to \mu$, primary muons and muons from tau decays cannot be distinguishable reliably in the final state. Hence, we will just add these contributions and leave only muons. But if tauons decay hadronically, we do consider them in final states, which is denoted by τ_h.

In summary, we consider the following distinct final states: 3μ, 4μ, $m\mu + n\tau_h$ ($m + n > 2, n \neq 0$), $2\tau_h + \not{E}_T$, $\mu\tau_h + \not{E}_T$, and $2\mu + \not{E}_T$. The simulated events are generated by `MadGraph` [25], then we make use of the `CheckMATE` framework [24] to encode a set of LHC analyses with final states we are interested in. Except for a few relevant analyses which have already included in `CheckMATE`, we encoded a total of 281 new signal regions defined in 28 different papers. We find that the dedicated Z' search base on 4μ final states is indeed the most sensitive one for mass range of 10 GeV $\leq m_{Z'} \leq$ 60 GeV; however, for larger masses, analyses containing three charged leptons final states are more sensitive.

3.3 Application to LHC Data

Based on above discussion, we encode the $L_\mu - L_\tau$ model by `FeynRules` [69] and output a `UFO` format model file. It is used by `MadGraph` in order to generate parton-level events. By taking advantage of `MadGraph`'s syntax, we define multiparticles `mt-` as charged leptons (μ^- and τ^-), and `miss` as invisible particles (μ and τ neutrinos, antineutrinos, and DM particles). Then it's pretty easy to generate different signals by this definition. For example, the $2l$ signal events can be generated by

specifying events containing two opposite sign `mt` and two `miss`; for the $3l$ signal, which is three `mt` plus a `miss`; and finally the $4l$ signal is just two pairs of opposite sign `mt`. All these signal events must contain a Z' mediator.

It is also important to note that, the Z' is allowed to be off-shell, but we did not include the interference between Z' and Z/γ. These interference terms vanish in the narrow width approximation, i.e. for $\Gamma_{Z'} \to 0$. One might question that, in the high mass region $g_{\mu\tau}$ is less constrained and the upper limit is large, which will lead to sizable interference terms. However, in our study which respect the perturbativity constraint in Eq. (3.16), the interference accounts for at most 6% of the squared Z' exchange contribution after cuts. This is considerably less than the typical effect of QCD NLO corrections, therefore we ignore these contributions. Moreover, in the range of $m_{Z'} > 100$ GeV, the interference terms are positive, so it is also conservative to ignore them.

While these signals events are generated by `MadGraph`, they are further passed to `Pythia 8.2` [26] for parton showering and hadronization. This step can also be done within `MadGraph` since `Pythia` is also inherited in `MadGraph`. After these preparation steps, we pass them to `CheckMATE 2` which applies the selection cuts based on the designated search regions, then it outputs the decision whether a given model is excluded or not. The search regions are defined based on LHC analyses, and we have encoded a total of 281 new signal regions which are not included in `CheckMATE`. The details can be found in Table 3.1.

List of Analyses	Center–of–mass energy		
Topologies	7 TeV	8 TeV	13 TeV
$2\mu + \not{E}_T$	[70, 71]	[70, 72–74]	[75–85]
$(2\tau_h \text{ or } \mu\tau_h) + \not{E}_T$			[86–89]
$3\mu \text{ or } 4\mu$		[90]	[65, 79, 80, 84, 85, 91–97]
$m\mu + n\tau_h$ $(m+n>2,\ n\neq 0)$		[98]	[79, 92, 96, 97]

Table 3.1: All analyses used in this study.

In the rest of this chapter, we will give detailed analyses of different signal searches containing $2l$, $3l$ and $4l$ final states. Just a reminder, here l only includes muons and hadronically decaying tauons. The analyses we encoded cover center of energy $\sqrt{s}$ at 7, 8 and 13 TeV. But in practice, analyses at 13 TeV are always more constraining, part of the reason could be that they were published more recently. Moreover, a general finding in all searches is that, in the final state, if we replace a muon by a hadronically decaying tauon, the sensitivity always decreases. Even though the branching ratio for tauon decaying hadronically is pretty large, about 65%, the τ-tagging efficiency is well below muon identifying efficiency. Therefore, it is more likely to misidentify QCD jets as hadronically decaying tauons. But we do consider contributions of τ leptons to the final sensitivity, due to the decays of $\tau \to \mu$.

The constraints on $L_\mu - L_\tau$ gauge boson were performed in two distinct scenarios. In the first case, we assume that Z' does not decay into dark matter particles. This can be done by either setting $q_{DM} = 0$ or $m_{DM} > m_{Z'}/2$ or both. Then there are only two parameters left affecting the strengths of signals: the mass $m_{Z'}$ and the coupling $g_{\mu\tau}$. For each distinct parameter setting, we generate at least 20,000 events, and if the Monte Carlo statistical error is dominated at the total error, we will

generate more events until having a stable result. Since the signal rates are proportional to $g^2_{\mu\tau}$, we are therefore able to determine the upper bound using the analyses results of a given $m_{Z'}$, which can typically be done in three or four trials.

The upper bound of coupling $g_{\mu\tau}$ (upper) and the corresponding cross section (lower) are shown in Fig. 3.2. Note that the cross sections include Z' decays, but we treat each τ as a charged lepton irrespective of its decay. The curves terminate in the region of large Z' mass when the perturbativity limit Eq. (3.16) is reached. The upper frame is the upper limits we obtained by `CheckMATE` and other existing constraints on $g_{\mu\tau}$ as functions of $m_{Z'}$. It includes separate bounds from analyses of final states with two, three and four charged leptons, which are plotted in green dot-dashed, red dashed and dark blue solid curve correspondingly. For $2l$ final states, we ignore mass region of $m_{Z'} < 10$ GeV, since the cut efficiency is quite poor here. `CheckMATE` cannot make a stable prediction, so we need to generate much more events in order to obtain reliable results. Since the resulting bound from $2l$ is clearly worse than $3l$ and $4l$ analyses, surely that is also the case in low mass region, then we save our efforts for $2l$ analyses.

As we can see, the curves in the left frame are not always smooth, for example in the region of 40 GeV $< m_{Z'} <$ 100 GeV at $3l$ final states (red dashed curve). This is because that, by going through all the signal regions of different analyses, `CheckMATE` will use the one with best expected sensitivity to set the bounds. This means that, for different Z' mass, the best signal region given by `CheckMATE` might change, which leads to discontinuities. The good reason is it avoids "look elsewhere" effects. The shaded area in Fig. 3.2 shows the value of $g_{\mu\tau}$ reproduces the measured $g_\mu - 2$, which includes Z' exchange. The anomaly value we use is taken from [99].

$$\delta a_\mu = a_\mu^{exp} - a_\mu^{th} = (29.0 \pm 9.0) \times 10^{-10} \,. \tag{3.17}$$

The brown solid line corresponds to the central value in the above equation; the darker and lighter shaded regions are corresponding allowed regions to 1 and 2 standard deviations.

At last, the cyan dot-dashed line represents the non-LHC bounds, which is derived from a measurement of neutrino “trident” production by the CCFR [100] ($m_{Z'} > 4$ GeV), and 4μ searches by the BaBar [101] ($m_{Z'} < 4$ GeV). In the case of CCFR, the upper limit is obtained by CL_S method with a 95% c.l.; which is the same as `CheckMATE` did. The upper bounds derived in this way is about 20% weaker than in [102], which seems to be based on the upper bound of the central CCFR cross section value plus 1.64 times the error. However, it is interesting to notice that, by averaging the cross section which are measured by CCFR and CHARM–II [103], we can get a similar bound as in CL_S method. On the other hand, in the case of BaBar, we smoothed the actual bound, since it fluctuates rapidly (about ±30%) around the smoothed-out line. These two experiments are shown to be the most stringent in the parameter space we considered ($m_{Z'} < 500$ GeV), therefore we do not include bounds from other experiments in Fig. 3.2, such as tests of lepton universality [104].

As shown in Fig. 3.2, the non-LHC bound is generally the most constraining one, except for the range of 10 GeV $< m_{Z'} <$ 60 GeV. In this region, the most stringent constraint comes from a published LHC analysis [65] that specifically designed to search for the $L_\mu - L_\tau$ gauge boson, which is based on $Z \to 4\mu$ decays at mass region 5 GeV $< m_{Z'} <$ 70 GeV in the CMS detector. However, for a given $m_{Z'}$, the constraints we obtained by `CheckMATE` is slightly weaker than the LHC analysis. This presumably comes from the different detector simulating strategies employed by of the `Delphes 3` [27] and `Geant 4` [105]. The former used a fast simulation while the latter taken by CMS collaboration used a full simulation.

However, for Z' mass outside the range 10 GeV $< m_{Z'} <$ 60 GeV, the best LHC bounds are based on other searching processes. In particular, the search for 4μ final states including softer muons as shown in [97], is comparable to or sometimes stronger than [65] at mass range of $m_{Z'} < 10$ GeV. On the other hand, for $m_{Z'} > 60$ GeV the tightest LHC constraints always comes from 3μ searches; the important ones are [91], [97] and [79]. This mainly results from the difference between $3l$ and $4l$ cross sections. To be more specific, the cross section of $3l$ final state, which is mediated by charged W-bosons (the rightmost diagram in Fig. 3.1), is enlarged by a factor of $2.5 - 3$ as compared to the $4l$ final state mediated by neutral Z/γ (the leftmost diagram in Fig. 3.1). Besides, the analyses with high sensitivity to $3l$ searches shows a little better cut efficiency than $4l$ searches.

As we mentioned above, the non-LHC bound is generally the best; only one dedicated LHC analysis [65] outperforms it at the designated mass region. This indicates that, if more analyses are designed to search for signals of $L_\mu - L_\tau$ model, the overall LHC bound might be better than the existing non-LHC bound. Moreover, with the increase of $m_{Z'}$, the upper bound on $3l$-signal cross section turns to be flat or slightly increase, as shown in the right frame of Fig. 3.2. This further proves that the current cuts were not optimized for $L_\mu - L_\tau$ model. An example is in [97]; the sensitivity decreased because of a transverse mass cut at larger mass region. The optimization on $L_\mu - L_\tau$ model is a main part in this thesis, and will be discussed thoroughly in next chapter.

So far we have discussed the first scenario of constraining the $L_\mu - L_\tau$ gauge boson, where DM particles are prohibited in on-shell Z' decays. If instead we turn it on, the branching ratio for $Z' \to l^+ l^-$ will decrease, which leads to a reduction on the cross section of $3l$ and $4l$ final states. Of course, this will also decrease the derived upper bound on $g_{\mu\tau}$. However, the branching ratio of $Z' \to \phi^*_{DM}\phi_{DM}$ is at most 25% for scalar DM particles, even in the case of $q_{DM} = 2$. Therefore, the upper bound derived from $3l$ and $4l$ final states is at most reduced by a factor of $\sqrt{3/4} = 0.86$.

However, the situation for $2l$ final states is more complicated. In this case, the signals come from both left and middle diagrams in Fig. 3.1, where left Z' decays invisibly and middle Z' decays into 2 leptons. If we allow $Z' \to \phi_{DM}\phi^\dagger_{DM}$ decays, then the contribution from left diagram increases, but from middle diagram decreases. This combination leads to a unclear prediction to final $2l$ signals. Even though the branching ratio of invisible Z' decays is no more than 50% in the parameter space we considered, the total cross sections cut efficiencies of these two diagrams are different. Therefore, we also need to perform a numerical analysis on $2l$-channels.

The results of constraints on $2l$ final states including dark matter particles are shown in Fig. 3.3. The upper frame is with charge $q_{DM} = 1$ and the lower with 2. In the gray region below the diagonal of plots, DM mass is larger than a half of Z' mass; hence it can only decay off-shell, which gives much smaller contribution compared to the on-shell decay. In this case, it goes back to the scenario without DM particles in Fig. 3.2. The green regions are excluded by our recast of analyses of 2μ final states. Besides, the horizontal black lines are corresponding exclusion limits in the absence of $Z' \to$ DM. As we can see in both frames, the green regions go beyond the black lines. This shows that, by adding DM particles into $2l$-channels, the sensitivity increases; and a larger charge ($q_{DM} = 2$) has a slightly better effect. However, the entire green region is already excluded in the case of $3l$ and $4l$ final states, which implies that in this model the production of DM particles are not as sensitive as leptons at LHC. Also note that, in our numerical study, the best bounds come from analyses [78, 80, 83]; and the cut efficiencies for the left and middle diagrams in Fig. 3.1 are indeed different.

So far the DM candidate we have considered is only limited to scalar particles, it can also be extended into spinor particles, such as a Dirac fermion χ. But the spin of invisibly particles which produced by on-shell Z' decays cannot be determined experimentally; the only quantity which makes

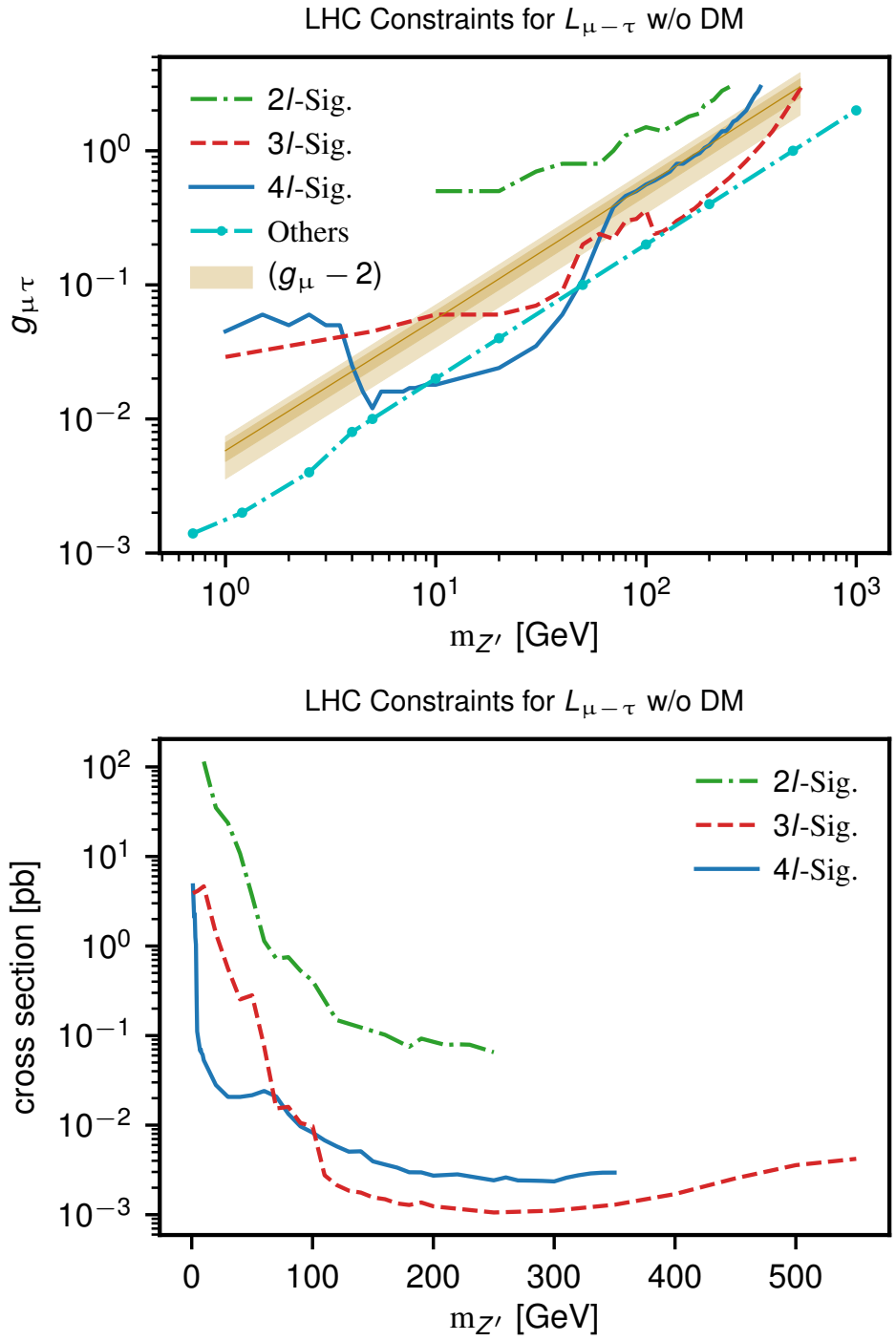

Figure 3.2: The upper limit of new coupling $g_{\mu\tau}$ (upper) and the corresponding cross section (lower). The shaded region in the upper frame shows the value obtained by the $g_\mu - 2$ measurement; the lower cyan dot-dashed curve is a summary of existing non-LHC constraints.

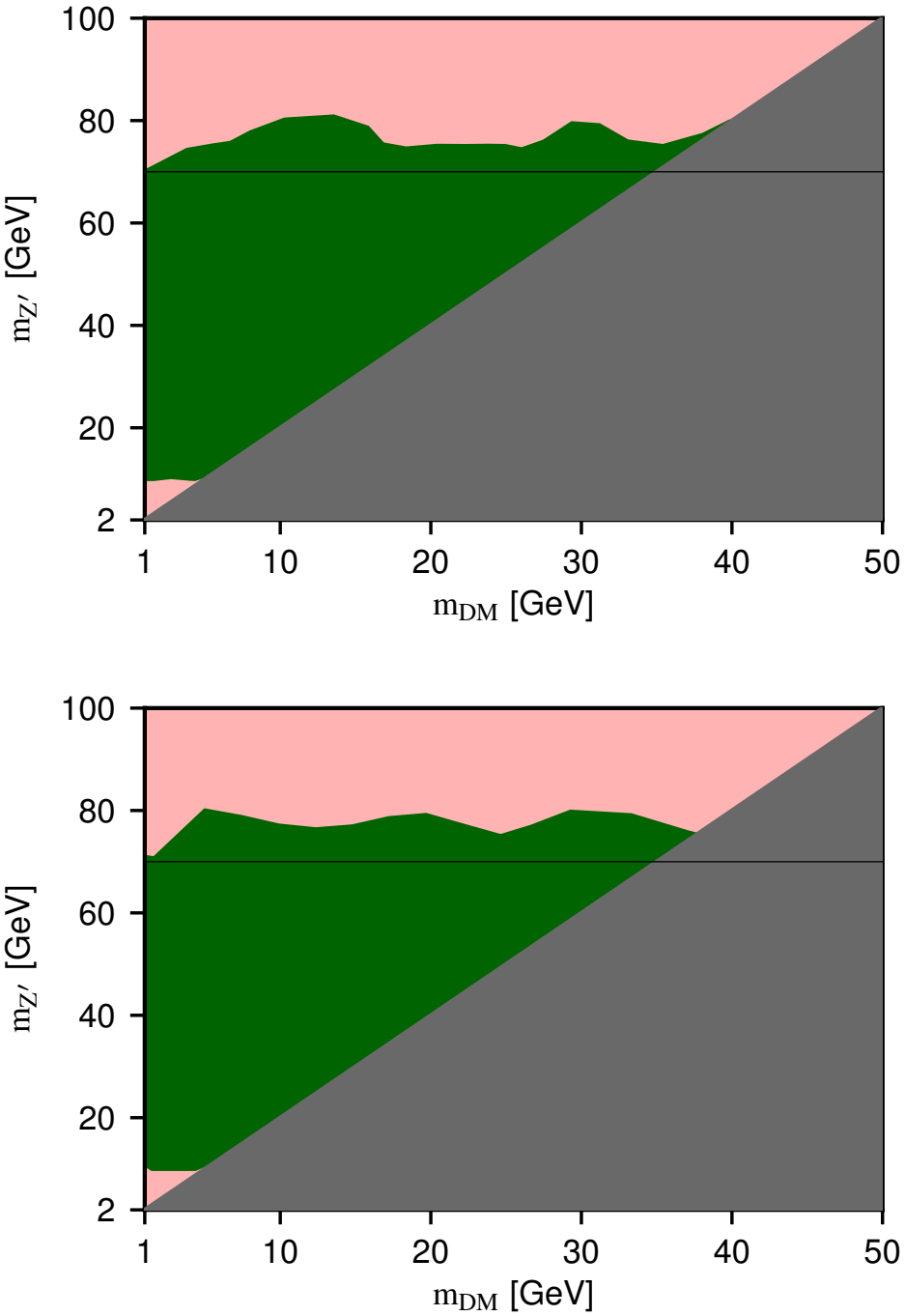

Figure 3.3: The constraints from Z' decays into final states of $2l$ and dark matter particles with coupling $g_{\mu\tau} = 1$ and $q_{DM} = 1$ (upper) and 2 (lower). The gray area indicates the parameter region that the corresponding DM decays are kinematically forbidden. The green region is excluded by analyses of 2μ final states; the black line represents the limit in the absence of decays into DM. The pink region which includes the green region, is excluded by analyses of 4μ final states.

a difference is the branching ratio of invisibly Z' decays. For Dirac fermion χ, the corresponding partial width is

$$\Gamma(Z' \to \bar{\chi}\chi) = \frac{m_{Z'}}{12\pi}\sqrt{(1-4z_{DM})}(g_V^2 + g_A^2 + 2z_{DM}(g_V^2 - 2g_A^2)) \,. \tag{3.18}$$

where g_A is the axial vector coupling, g_V is the vector coupling, and $z_{DM} = m_\chi^2/m_{Z'}^2$. If $g_V = 0$ and $g_A = g_{\mu\tau}$, it is the same as the scalar DM with $q_{DM} = 2$, which is shown in the lower frame of Fig. 3.3. On the other hand, if $g_A = 0$ and $g_V = g_{\mu\tau}$, the partial width is larger than scalar DM while the mass m_{DM} increases. But the derived bounds would still be weaker than the case of $3l$ and $4l$ final states, since the branching ratio of $Z' \to$ DM won't exceed 25%.

3.4 Conclusions

In this chapter, we have discussed the anomaly free gauged $U(1)$ extension models of SM, in particular $U(1)_{L_\mu - L_\tau}$ model. It has no coupling with electrons, which makes it evade from experimental searches like LEP. To further study its constraints in LHC experiments like ATLAS and CMS, we recast a large number of LHC analyses taking use of the framework of `CheckMATE`, which are summarized in Table 3.1. Here we simplify the model by restricting its parameters in Z' mass and the coupling $g_{\mu\tau}$. The signal final states we considered are $2l$, $3l$ and $4l$ states with Z' masses up to 550 Gev to respect the perturbativity condition; here l means a muon or hadronically decaying tauon. In principle, $2l$ final states should have the highest sensitivity, since the decays of $Z' \to$ DM. However, our study shows that LHC data are not sensitive enough to the production of DM particles in this model; and the bound on $g_{\mu\tau}$ derived from $2l$ is always weaker than $3l$ and $4l$ final states. Moreover, if we replace a muon by a hadronically tauon in the final state, the sensitive always decreases. In the range between 5 GeV $< m_{Z'} <$ 60 GeV, the best LHC bound comes from 4μ final states, otherwise from 3μ final states. We also include bounds derived from low energy non-LHC experiments; in most cases, they are shown to be the best, except for 10 GeV $< m_{Z'} <$ 60 GeV. In this region, the strongest constraint comes from a CMS analyses [65].

This is the only analysis which is specifically designed to look for Z' bosons. Therefore, it is very likely to improve the LHC bounds by optimizing cuts in the rest of Z' mass region. Moreover, this analysis is based on searches for $Z' \to 4\mu$; but we have shown that, in our study, $3l$ final states are more sensitive than $4l$ states for larger Z' masses, which further indicates an optimization is possible. On the other hand, by statistically combining final states with muons and with hadronically decaying tauons, it is also possible to improve the current sensitivity, since we can derive the relative normalization of these channels in this model. A further study of optimization on this model will be discussed in next chapter using Machine Learning methods.

CHAPTER 4

Optimize Cuts on $U(1)_{L_\mu - L_\tau}$ by Machine Learning Methods

Extending the Standard Model (SM) by a gauged $U(1)_{L_\mu - L_\tau}$ group [106] does not introduce new gauge anomalies even if we stick to the SM fermion content, but leads to potentially sizable positive contributions to the anomalous magnetic moment of the muon $(g_\mu - 2)$, whose SM prediction is too low by about $3.5\,\sigma$ [107]. Once right–handed neutrinos are introduced it also allows the construction of realistic neutrino mass matrices [108, 109], and can be used to construct realistic models of particle Dark Matter (DM) [62]. Moreover, since the model does not introduce extra couplings of the electron, the model avoids the strong constraints from $e^+e^- \to \mu^+\mu^-$ or $e^+e^- \to \tau^+\tau^-$ in e^+e^- collision experiments.

In Chapter 3 we studied [110] to what extent published LHC analyses can be used to constrain this model through the production and decay of the new Z' gauge boson. We also allowed for the existence DM particle charged under $U(1)_{L_\mu - L_\tau}$, either a complex scalar $(\phi_{\rm DM})$ or Dirac spinor $(\chi_{\rm DM})$. We found that for most values of the mass $m_{Z'}$ of the new gauge boson, published LHC analyses impose a weaker bound on the new gauge coupling $g_{\mu\tau}$ than non–LHC experiments, the latter being dominated by searches for low–mass Z' at BaBar [101] as well as neutrino "trident" experiments [100, 102, 103]. Only for 10 GeV $\leq m_{Z'} \leq$ 60 GeV does the best bound on $g_{\mu\tau}$ come from the LHC, thanks to a dedicated search by CMS [65] in the four muon final state with designate cuts specially for $U(1)_{L_\mu - L_\tau}$ extension of SM.

The sensitivity of LHC data can clearly be improved by applying selection rules that have been optimized to search for this specific Z' boson. In our previous analysis [110] we had seen that final states with muons always have better sensitivity than otherwise equivalent final states with hadronically decaying τ leptons. Moreover, since the cross section for production $\mu\nu_\mu Z'$ final states is considerably larger than that for $\mu^+\mu^- Z'$ production, the best sensitivity for Z' searches at the LHC is expected in the $3\mu + \not{E}_T$ final state, where $\not{E}_T$ stands for missing transverse energy. An exception may occur if the invisible width of the Z' is very large. The latter depends on the mass and charge of the DM particle, and can be probed in the $\mu^+\mu^- + \not{E}_T$ final state. In this chapter we therefore focus on these final states.

Since the new contribution to the $3\mu + \not{E}_T$ signal will be dominated by the production and decay of nearly on–shell Z' bosons, we design a simple set of cuts, assuming that one can guess the value of $m_{Z'}$ from the di–muon invariant mass distribution; this is essentially a classical "bump hunt". In addition, we develop and compare a variety of machine learning (ML) methods. Our goal is to

find a single classifier that has good sensitivity over a wide range of Z' masses, rather than devising dedicated searches for each value of $m_{Z'}$. Moreover, at least initially we consider a large number of input variables, including both low–level features (the 4–momenta of the final state objects) and higher–level features (e.g. invariant masses of pairs of final state objects); the latter are taken from published experimental analyses of multi–lepton final states.

We find that, after sufficient training, a deep neural net (NN) outperformed the other classifiers in most cases; in particular, it did considerably better than the simple bump hunt. However, a certain boosted decision tree (BDT) was nearly as efficient in separating signal from SM background. Moreover, it allowed to identify the most important input features, which helps to extract physical information from the ML algorithm; in contrast, the NN is basically a "black box". The information of the original BDT on feature importance also led us to devise simpler classifiers, both NN and BDT, with significantly fewer input quantities but equally good performance. Performing both a NN and a BDT analysis therefore guarantees excellent sensitivity of the Z' search, and physical insight in the important kinematical features of the signal.

The remainder of this chapter is organized as follows. In Section 4.1 we briefly describe the SM extended with the $U(1)_{L_\mu - L_\tau}$, focusing on different kinds of possible signal final states. In Section 4.2, we describe the data set and training process for ML based classifiers. In Section 4.3, we discuss the results from this new approach. Finally, in Section 4.4 we summarize our study and draw some conclusions.

4.1 Model and Signal

We have discussed the theory of $U(1)_{L_\mu - L_\tau}$ gauge extension model in previous chapter. As a remainder, in this section we will briefly recap the formulism and notations. To extend the SM gauge group by local $U(1)_{L_\mu - L_\tau}$ symmetry requires introduction of a new gauge boson Z', which can also be a mediator connecting SM to DM particles; the corresponding field strength tensor is $\mathcal{Z}'_{\mu\nu} \equiv \partial_\mu Z'_\nu - \partial_\nu Z'_\mu$, while the covariant derivative instead of the normal partial derivative can be used to describe the interactions, i.e. $\partial_\mu \to D_\mu = \partial_\mu - i g_{\mu\tau} q_{\mu\tau} Z'_\mu$, where $g_{\mu\tau}$ is the new gauge coupling and $q_{\mu\tau}$ the corresponding $L_\mu - L_\tau$ charge. The model may contain a complex scalar DM particle $\phi_{\rm DM}$ or a spinor DM particle $\chi_{\rm DM}$, which are singlets under the gauge groups of the SM but carry $L_\mu - L_\tau$ charge $q_{\rm DM}$. The DM particle affects LHC physics basically only through Z' decays into invisible final states.[1] As long as we keep the mass and charge of the DM particle as free parameters, we can therefore fix its spin without loss of generality. For definiteness we consider the scalar DM particle here.

The LHC signals we consider stem from the production and decay of (nearly) on-shell Z' bosons. The above assumptions about the particle spectrum imply that at leading order the Z' can only decay into second or third generation leptons, and possibly into DM particles. The corresponding partial width are shown in eq.(3.14). Our perturbative analysis will not be reliable if the new gauge coupling is very large. We therefore only consider scenarios where the total Z' width is smaller than $m_{Z'}$, which implies eq.(3.16) holds.

The example Feynman diagrams of signals that we are interested in are shown in Fig. 3.1: the Z' can be emitted off a $\mu^+\mu^-$ or $\tau^+\tau^-$ pair ; off a second or third generation $\nu\bar{\nu}$ pair; and off a $\mu\nu_\mu$ or

[1] The cross section for producing DM particles via the exchange of a virtual Z' is much smaller than that for producing an on–shell Z' decaying into neutrinos, and will thus have negligible impact on the final state we consider here.

$\tau\nu_\tau$ line. Our assumptions imply that the only visible particles that can be produced in Z' decays are muons and tau leptons. Invisible Z' decays in the left figure and visible Z' decays in the middle contribute to the $2l$ signal; recall that the former can receive contributions from Z' decays into DM particles. Visible Z' decays in the left diagram leads to $4l$ signal; the CMS analysis [65] investigated this final state for the case that all leptons are muons which originate from the decay of an (almost) on–shell Z boson, offering good sensitivity for 10 GeV $\leq m_{Z'} \leq$ 60 GeV.

In the right figure, invisible Z' decays lead to single lepton final states, which we do not consider because of the very large SM background from the production of (possibly off–shell) leptonically decaying $W^\pm$ bosons. Visible Z' decays here lead to $3l$ signals. Note that this class of diagrams offers a significantly larger cross section (after summing over both possible charges) than those giving rise to $4l$ final states. In our previous work [110] we indeed found the best sensitivity for $3l$ final states, except for the mass range that can be probed in the decay of on–shell Z bosons [65].

In the above discussion l stands for a μ or τ lepton. The former are stable as far as the LHC experiments are concerned, and are straightforward to identify experimentally, if they are produced sufficiently centrally and with sufficient transverse momentum p_T (the precise requirements will be given below). In contrast, tau leptons decay very quickly. $\tau \to \mu\nu_\mu\nu_\tau$ decays contribute another, softer, muon to the final state. $\tau \to e\nu_e\nu_\tau$ decays lead to qualitatively different final states, which come with their own sources of background. Since our Z' does not couple to electrons, replacing a muon (pair) in a multi–muon final state by an electron (pair) will greatly reduce the signal cross section, whereas the SM background, being essentially flavor universal, will remain the same; this therefore results in a final state with much worse signal to background ratio. Finally, τ leptons can decay into hadrons plus a ν_τ; however, these decays are not easy to identify experimentally, and suffer from considerably additional backgrounds. The upshot of this discussion is that we expect the best sensitivity in final states defined exclusively via the number of muons and missing E_T; indeed, this is what we saw in our previous study [110]. It should be noted that the 3μ signal also receives a (small) contribution from the left diagram of Fig. 3.1 if at least two of the leptons are τ's, one of which decays into a muon while the others decay hadronically. Similarly, the right diagram can contribute to the 2μ final state.

In order to simulate the 2μ and 3μ backgrounds and signals we use `MadGraph` to generate the process $pp \to m\mu + n\tau + (4-m-n)\not{p}$, where $\not{p}$ means neutrinos or DM, $m, n \geq 0$, and $3(2) \leq m+n \leq 4$, under the condition that only events with exactly two or exactly three muons in the final state are accepted; the signal contribution is defined by requiring at least (in practice, exactly) one Z' propagator in the Feynman diagram, as shown in Fig. 3.1.

4.2 Machine Learning based Methods

We use gradient boosting decision tree (GBDT) and deep learning neural network as tools to discriminate the possible new physics signals and the Standard Model background. NN is less prone to be affected by the choice of features. It is shown that, NN can perform well even with very basic features [111]. But it is hard to explain as a dark box. However, GBDT is more instructive, as we can get the relative feature importance of our model, which means, we know which feature is more important when we actually design the cut in event selection.

The basic pipeline is, first, for each possible value of $m_{Z'}$ (10, 50, 100, 200, 300, 400, and 500 GeV), we generate 1 million signal and 1 million background events. we change the random seed for

the generation of background events to ensure that the final data we use for training are statistically independent. We use a single classifier instead of individual one for every single mzp, because we want to check how well the ML algorithms are able to understand the mixed data. In the real application to detect new physics BSM, we need a powerful classifier not only for various settings of variables in the same model, but also for distinct models. Then, after some pre–selection which we will discuss later, we use `CheckMATE` [23, 24] to extract all the features we need for the training process. `CheckMATE` is a package, which defines many observables in the detector, based on many packages including [27, 28, 112–121]. Next, we combine all these events together, including different value of $m_{Z'}$, and use machine learning to train a classifier. Finally, we get our bound for $g_{\mu\tau}$ by the use of background rejection and signal efficiency of this classifier. The bound here means the sensitivity, not the real bound from the experimental data. To calculate the sensitivity, we assume a Poisson distribution centered at the events number of background, and the possibility of observing higher event numbers than the number of signal plus background should be at least 5%. Otherwise, the signal is excluded. The event number we use here is normalized from the simulated event number to the real one in published ATLAS and CMS 13 TeV paper, which has integrated luminosity equaling to $36.1/fb$.

In our simulation, events are generated by the Monte Carlo generator `MadGraph` [25], hadronized by `Pythia` [26], and simulated the detector response by `Delphes` [27]. The NN is implemented by the framework `Keras` [2] and `TensorFlow` [3], and GBDT by `XGBoost` [4].

4.2.1 Features

The primitive events cannot be used directly for training, we need to extract potential low–level and high–level features [111] with the help of `CheckMATE`. Low–level features refer to some basic variables measured in detector, like four momentum. And high–level features are some hand–designing variables, like invariant mass of di–lepton pair. To be more general and less human intervention, we just take all the common variables appeared in LHC analyses, as shown in Table 4.1. All the momentum and energy in the feature list with label `i` are ranked in the descending order of transverse momentum (p_T), and hence the leading one has the largest p_T.

In order to get well defined final state objects, we need do a pre–selection first. In detail, we only consider final electrons have a minimum p_T of 10 GeV and $|\eta| < 2.5$, final muons with $p_T > 10$ GeV and $|\eta| < 2.4$, final jets with $p_T > 25$ GeV and $|\eta| < 2.4$. Moreover, we require jets separated from any lepton by $\Delta R > 0.4$. While calculating these features, we only record variables related to muon, jet or missing energy, and ignore electron, as it does not couple to Z' boson. $m_{\mu^+\mu^-}$ is calculated by the only opposite sign di–muon pair for 2μ–signal, while for 3μ–signal, $m^{(1)}_{\mu^+\mu^-}$ ($m^{(2)}_{\mu^+\mu^-}$) which have the invariant mass closest to (away from) Z boson mass. Then we use the same di–muon pair to get the stransverse mass m_{T2} for 2μ–signal, and $m_{T2}^{(1)/(2)}$ for 3μ–signal. Finally, we exclude events with $\not{E}_T < 10$ GeV or $\not{E}_T < 100$ GeV. These two values are empirical, but it can improve the performance of our model, and might also reduce outliers in samples.

[2] https://keras.io/
[3] https://www.tensorflow.org/
[4] https://xgboost.ai/

Features	Definition
p_i	Four momentum of leptons and jets
ϕ_i	Azimuthal angle of leptons and jets
η_i	Pseudorapidity of leptons and jets
p_T^i	Transverse momentum of leptons and jets
$\not{E}_T$	Missing transverse momentum
m_T^i	Transverse mass [24] of leptons and jets
$m_{\mu^+\mu^-}$	Invariant mass of the muon pair for 2μ–signal
$m^{(1)}_{\mu^+\mu^-}$	Invariant mass of the muon pair which is closest to m_Z for 3μ–signal
$m^{(2)}_{\mu^+\mu^-}$	Invariant mass of the muon pair which is different with $m^{(1)}_{\mu^+\mu^-}$ for 3μ–signal
m_{T2}	Stransverse mass [117], calculated by $m_{\mu^+\mu^-}$ for 2μ–signal
$m^{(1)}_{T2}$	Stransverse mass [117], calculated by $m^{(1)}_{\mu^+\mu^-}$ for 3μ–signal
$m^{(2)}_{T2}$	Stransverse mass [117], calculated by $m^{(2)}_{\mu^+\mu^-}$ for 3μ–signal
$\not{E}_T / H_T$	H_T is the scalar sum of p_T of leptons and jets

Table 4.1: List of features we used for training.

4.2.2 Machine learning

After the above pre–selection, we get a data set with a total number of 2, 500, 000 (710, 000) samples in case of $\not{E}_T > 10$ GeV ($\not{E}_T > 100$ GeV), with an average ratio of background 0.51 (0.14). Then we randomly select 90% subset for train, 10% for test. The preprocessing step is simple, we take the absolute value of angle related features, like η and ϕ, and standardize all features by removing their mean and scaling to unit variance. The neural network is a simple fully connected network with linear layers. It consists of five linear hidden layer, all using `relu` as activation function except the last output layer, which is a sigmoid function. We also add two dropout layer with dropout ratio 0.1 to reduce overfitting. We use Adam optimizer with a learning rate of 0.0002 to update model weights. The training process is based on a mini–batch of size 64 with a maximum epochs of 50. For XGBoost model, we use maximum 1500 estimators, maximum depth 11, a fraction of subsampling features 0.8, and a learning rate 0.01. For both neural network and XGBoost, early stopping is used to prevent overfitting.

Since the signal to background ratio varies among values of $m_{Z'}$, a simple metric of accuracy is not enough to evaluate model performance. Instead, we use the area under the receiver operating characteristic (ROC) curve, or simply the area under curve (AUC), as our metric. ROC curve is a plot of the true positive rate against false positive rate at all possible thresholds. Therefore, a larger AUC, means the model generally performs better, and it is threshold–invariant. In Fig 4.1, we show the ROC curve of XGBoost and NN. In test set of $\not{E}_T > 10$ ($\not{E}_T > 100$), the AUC score is 0.9638 (0.9831) for XGBoost, and 0.9605 (0.9819) for NN. Therefore, in our case, XGBoost slightly outperforms NN. But that could be because of the notorious hard training process of NN. Additionally, since these features could potentially relate with each other, like mass and four–momenta, our features are similar to the case in natural language processing (NLP). In NLP task, such as neural machine translation (NMT), the meaning of a word is dependent on other words in the sentence. So we could treat features

of one sample as a sentence, and use recurrent neural network (like LSTM) or one–dimensional convolutional neural network to model it. But, unfortunately, they perform very much the same as the simple fully–connected neural network (fc NN). So, for the sake of simplicity, we pick XGBoost and fc NN as our final models.

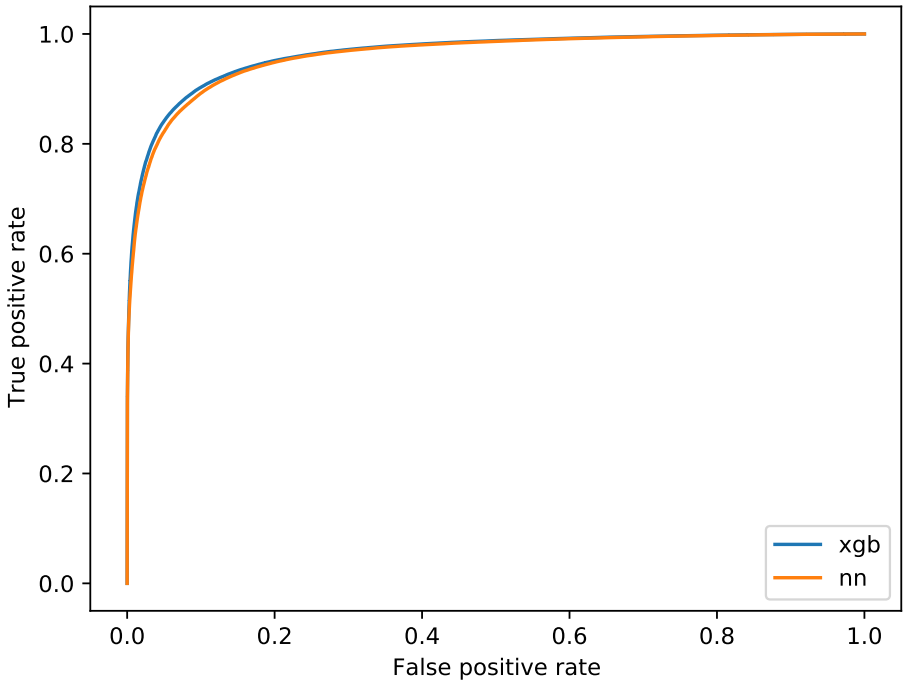

Figure 4.1: ROC curve of XGBoost and NN.

For comparison, we also use a dedicated cut for different values of $m_{Z'}$ separately as references. The cut works as follows: we first put a cut at $|m^{(1)}_{\mu^+\mu^-} - 91.19| = 8$, any event with $|m^{(1)}_{\mu^+\mu^-} - 91.19| < 8$ is excluded as background; then, for the remaining events, we take events in the region of $0.9\, m_{Z'} < m^{(Z')}_{\mu^+\mu^-} < 1.1\, m_{Z'}$ as signal, where $m^{(Z')}_{\mu^+\mu^-}$ means the mass of the muon pair that is nearest to $m_{Z'}$, otherwise background. The detailed results will be shown in next section.

4.3 Application to the Phenomenology at LHC

In this section, we apply the method described in previous section with NN and XGBoost respectively. From the combination of NN and XGBoost, we read the new expected upper limits of interaction couplings with selected variables, while the XGBoost tells that the most noticeable features that can discriminate the signal and background. The classifier itself contains the physical information inside the event signatures at colliders. We will show how the information from feature importance of GBDT helps us understand physical properties.

In order to show our model actually works at physical level, not just some magic black box, we use a similar method as in Ref. [111], which is to compare the normalized event distribution of simulated signal, background and model predicted signal at different features. Since model prediction depends on threshold, here, we choose it to be the value when our model hits a precision (background rejection) of 0.90. Therefore, any event with a possibility larger than this threshold will be considered as a predicted signal. The result is shown in Fig. 4.13 4.14 4.15 4.16 4.17 4.18 4.19, blue line is pure simulated SM background, orange line is pure simulated signal, and green line is our model predicted signal selected among a mix of these simulated background and signal. Additionally, all these events are from the untouched test data, so it means our model could actually discriminate signal from background among unseen data.

4.3.1 3μ–Signal in LHC without DM Phenomenology

When constraining $g_{\mu\tau}$ without DM in final state in our previous research, $3l$–signal performs better than $4l$–signal in general, as the larger cross section. The only exception is the mass window from 10 GeV to 60 GeV, as the selection rules designed by CMS collaboration. We, therefore, hope the final results should be better than $4l$–signal in all range, since the $3l$–signal is stronger, and the ML classifiers are expected to offer better classification than pure cut based selection. In the previous research, we use $3l$–signal, while we use 3μ–signal this time. The new result compared to the bounds from LHC and non–LHC analyses is shown in Fig. 4.2 and Fig. 4.3. Although we assumed the result is from $3l$–signal instead of 3μ–signal, when we recast published analysis, the best constraints are actually from 3μ–signal. Therefore, we are able to compare the old result with the recent ML algorithms directly. Fig. 4.2 shows the best results from our classifier with extra validation data generated for $m_{Z'}$ = 15, 25, 75, 150, 250, 350, and 450 GeV to check whether our algorithms work for the parameter settings that never appear in training set, while Fig. 4.3 shows more detailed results for different selection and preselection rules. In Fig. 4.3, the result from our ML based classifier with $\not{E}_T > 10$ GeV is better for $m_{Z'} < 100$ GeV, while $\not{E}_T > 100$ GeV wins for $m_{Z'} > 100$ GeV. It means that the event distributions for $m_{Z'} < 100$ GeV and $m_{Z'} > 100$ GeV in phase space are probably separated. In both cases, our classifier distinguish the signal and background quite well. To compare with the dedicated cut based analysis, we use simple cut 1 (SC1) and simple cut 2 (SC2) as comparison for every mass point. Since we've already known the dedicated cut around $m_{Z'}$ peak should be very effective, especially for the case that $m_{Z'}$ = 10, 50 GeV, as the histogram in Fig. 4.4. Our universal classifiers still win in most of points.

The bound from our ML based classifier has the similar behavior with cut based analysis, where the bound for more massive Z' turns weaker gradually. In our test, the result implies that we should always include some rules of pre–selection, since the model trained with $\not{E}_T > 1$ GeV performs much worse than the models with $\not{E}_T > 10$ GeV or $\not{E}_T > 100$ GeV. The main reason is that the generated data contain anomalies and outliers that could mislead the classifier during training process, and hence lead worse classification. The difference between $\not{E}_T > 10$ GeV and $\not{E}_T > 100$ GeV claims that the program can understand the data well for better pre–selection, when the pre–selection keeps the necessary information in the data, rather than discards them. For light mass points, the model with $\not{E}_T > 10$ GeV is better than the model with $\not{E}_T > 100$ GeV. In contrast, the model with $\not{E}_T > 10$ GeV has lower sensitivity for $m_{Z'} > 100$ GeV. The reason is that the events with $m_{Z'} < 100$ GeV in phase space are quite different with $m_{Z'} > 100$ GeV, and hence our universal classifier may be influenced. The $\not{E}_T > 100$ GeV removes more events with lighter mediator, and makes the classifier

perform better for heavy mediator. The universal classifier may lose sensitivity if the distribution of the events in phase space is not concentrated. If we increased the size of the data, especially in the less sensitive region, and carefully trained more complex network, we would have better universal classifier. Nevertheless, here we want to focus on whether more information can be read from our ML algorithms. We, therefore, firstly consider the feature importance for $\not{E}_T > 10$ GeV, since it contains more events after preselection, compared to $\not{E}_T > 100$ GeV to show more details about our algorithms.

We can read the detailed information from the figure of feature importance and histograms of variant features. In figure 4.8, the top 9 features are $m^{(1)}_{\mu^+\mu^-}$, $p^{\mu 0}_T$, $m^{(2)}_{T2}$, $m^{(2)}_{\mu^+\mu^-}$, $m^{\mu 0}_T$, $m^{(1)}_{T2}$, $p^{\mu 1}_T$, $m^{\mu 1}_T$, and $\not{E}_T$. Additionally, more than half of the features have negligible feature importance, which should be unhelpful. We, therefore, compare 4 different cases, top 9 features, top 6 features, top 3 features, and all features excluding top 9 features in figure 4.5. Obviously, NN is more robust than XGBoost, when we remove all top 9 features. Theoretically, even if we remove all high level features, the well trained neutral network can only use low level features (4–momenta of final state particle) to reconstruct higher level features. Therefore, the result from NN should not change a lot. In contrast, the XGBoost require high level features to offer more information. Furthermore, to prove the feature importance indeed reflect physical information, not just a series in black box, we only keep 3, 6 or 9 features to do the classification. It is clear to see that, in the left hand side of figure 4.5, both NN and XGBoost can reach the similar results compared to the one from all features through just top 3 features. The AUC with only 3 features are still nearly 0.95. This means our result is physical, which reflects the information of the distribution in collider. Although, in the feature importance, less than half of the features have sizable importance, when we include the high level features, the most important features are only relevant to the most energetic 3 muons and the missing momentum, and this result connect to our physical view straightforwardly. The difference between NN and XGBoost, is that NN can use low level features, e.g. 4 momenta of final state particles, to form high–level feature automatically, e.g. invariant mass of particle pair, transverse mass, etc., while XGBoost, cannot internally find out those features. In the previous research [111], they claimed the same conclusion that the first layer of NN should learn the high level feature automatically. Therefore, if removing features with low feature importance does not influence the final result in both NN and XGBoost, it means that the remaining features abstract the topology of the final states very well. These are 2 different models with separated training process. Nevertheless, we get the coincident result that the core features help us to classify the events.

To read more information from feature importance, we consider figure 4.8 for $\not{E}_T > 10$ GeV, figure 4.10 for the input without top 9 features, and figure 4.9 for $\not{E}_T > 100$ GeV from tree based algorithms, and compare the result with previous conclusion from cut based analyses from ATLAS and CMS. For $\not{E}_T > 100$ GeV, the top 9 features are $m^{(1)}_{\mu^+\mu^-}$, $m^{(2)}_{T2}$, $m^{(1)}_{T2}$, $p^{\mu 1}_T$, $m^{(2)}_{\mu^+\mu^-}$, $m^{\mu 1}_T$, $p^{\mu 0}_T$, $m^{\mu 2}_T$, and $m^{\mu 0}_T$. For the input without top 9 figures and $\not{E}_T > 10$ GeV, the top 9 features are $E^{\mu 0}$, $E^{\mu 1}$, $\eta^{\mu 0}$, $p^{\mu 1}_x$, $p^{\mu 1}_y$, $p^{\mu 0}_x$, $\eta^{\mu 1}$, $E_T\,/H_T$, and $p^{\mu 0}_y$. Obviously, when we have high level features they are more useful than low level features. When we only consider low level features, the most energetic objects, i.e. top 3 muons, are still the most important.

Our signature has 3μ and sizable missing energy in the final state, while jets are either τ–tagged or soft. Moreover, there should be 2μ with opposite charge, which form a pair, has invariant mass similar with the mass of new gauge boson Z'. We can clearly see this result from figure 4.4. Since we mix final state with exactly 3μ from exact μ or τ decaying to μ, the final result have not only one peak around $m_{Z'}$, but also a bump around 40 GeV, because of the mismatch. Through the result in

figure 4.3, the mass specialized SC2 are quit efficient, and even better than the universal ML classifier at $m_{Z'} = 10$ GeV. Therefore, it should be very important feature to distinguish background and signal, especially for the case that have $m_{Z'}$ away from the mass of Z boson. From the feature importance, this feature rank the topmost, which agrees our physical perspective. Moreover, in the analysis from ATLAS and CMS, the mass related features and the transverse momenta are normally better than 4-momenta, while the feature importance gives the same conclusion. The comparison tells us that the ML based classification connects to the physical information as well as the cut based analysis, although the training process is purely data–driven. In sum, even if we use ML based classifier, instead of cut based one, we can still understand the topologies at colliders from feature importance without losing physical information.

In details, we draw the histogram with leading features to check whether the signal is distinguished from background as well as we expected. In figure 4.13 we see the signal is well separated from background for $m_{Z'} = 10,\ 50,\ 200$ GeV. It means that $m^{(1)}_{\mu^+\mu^-}$ is a good feature to distinguish signal and background for $m_{Z'}$, even if just applying one cut. On the contrary, in figure 4.13 (bottom left), the signal strongly overlaps with background. It means that $m^{(1)}_{\mu^+\mu^-}$ is not able to distinguish signal and background in this case, since Z' boson is quite similar with Z boson. As for the bound we got in figure 4.3 (left), in general, the curve is smooth and the bound gradually turns weaker; the only exception is the region around $m_{Z'} = 91$ GeV, where the Z' mass is very near to the Z mass in SM. In other words, the signal looks similar to the background for our universal classifier, and is hard to be distinguished from each other. Since the leading feature is $m_{\mu^+\mu^-}$, the overlap in the histogram for $m_{Z'} = 100$ GeV implies that the classifier is supposed to partially lose sensitivity in this mass region. Fortunately, when we use extra cut $\not{E}_T > 100$ GeV, the classifier focus on the case for $m_{Z'} > 100$ GeV, and the sensitivity at $m_{Z'} = 100$ GeV is improved. The coincidence between the physical perspective and data driven methods implies that even if we have to use NN, which are trained independently with a separated algorithms, to achieve the best sensitivity; the feature importance from XGBoost is still very helpful. Therefore, the combination of different ML based classifier helps us to understand the physical process better than a single one.

In summary, NN generates the high level features to classify events but in a black box; hence we don't have to work out the high level features by ourselves. However, without the help of BDT we don't know how the features in the black box relates to the physical signatures in collider. The features with similar importance may not rank in the same series, when comparing the rank from feature importance in XGBoost and the rank of the influence by removing every single feature. Nevertheless, when the importance of two features is quite different, the feature with much higher importance is indeed more important. Additionally, if we focus on the NN and XGBoost in the left part of Fig. 4.5, the result from XGBoost is slightly better. It means that after removing low–level features, which are used by NN to form high level feature internally, the results from NN requires carefully tuning to reach the best performance; and BDTs are able to use the high level features more efficiently.

For additional information, We draw more histograms in Section 4.5 (Fig. 4.13 4.14 4.15 4.16 4.17 4.18 4.19) to compare the distribution of signal and background and to check whether the selected features are good enough for classification. In the upper part of Fig. 4.13, which describe the $m_{Z'} = 10$ and 50 GeV respectively, the shapes of signal and background are similar with the case of cut based analyses. A cut at the crossing point of signal and background may perform efficient as well. Moreover, since the signal and background are well separated, our classifier can give the precise estimation of signal. On the contrary, in the lower part of Fig. 4.13, the difference between signal and

background is unable to be described by a single cut. The feature $m^{(1)}_{\mu^+\mu^-}$ helps us to distinguish signal and background at $m_{Z'} = 200$ GeV, given that their distributions are not strongly overlapped with each other. Moreover, if the feature is not used in ML algorithms, we should use $|m^{(1)}_{\mu^+\mu^-} - m_Z|$ instead for cut based analysis. The BDT, however, can automatically judge how to use the information in $m^{(1)}_{\mu^+\mu^-}$. When the signal and background are quite similar at $m_{Z'} = 100$ GeV, our ML based classifier partially loses sensitivity, since the overlap is too large. In Fig. 4.17 and Fig. 4.18, the lower left figures, which describe $m_{Z'} = 100$ GeV, also show bad performance of estimation (difference between orange line and green line). This implies the same conclusion from $m^{(1)}_{\mu^+\mu^-}$ that our classifier may lose sensitivity at the region around Z mass. From other figures (Fig. 4.15 4.16 4.17 4.18 4.19), we can read the similar conclusion that although the feature cannot be used as a single cut, the sizable isolation of signal and background may make the feature useful in the ML based classification, and this is the reason why such features rank at top. From the distribution of missing energy (Fig. 4.19), the peaks of signal and background are overlapped, and hence are hard to use in cut based selection. It means that the cut $\not{E}_T > 100$ GeV only helps to prepare the data to focus on the properties of $m_{Z'} > 100$ GeV, not that it distinguish the signal from background directly. This is the reason why we name it "preselection". From the figure 4.19, obviously, after preselection $\not{E}_T > 100$ GeV, most of the signals for $m_{Z'} = 10$ or 50 GeV are removed, while the remaining data are mainly from $m_{Z'} > 100$ GeV, and hence our signal remaining in the data set looks similar compared to the case with $\not{E}_T > 10$ GeV.

4.3.2 2μ–Signal in LHC with DM Phenomenology

From the ATLAS and CMS analyses, we have shown [110] that the DM phenomenology cannot be probed, since the relevant parameter region has already been excluded by $3l$–signal. Moreover, the bound from $2l$–signal is much weaker than the result from $3l$–signal. Therefore, even if the extra coupling with DM may decrease the $3l$–signal and increase the $2l$–signal, it is hardly to be probed. Nevertheless, the performance of analyses from ATLAS and CMS are achieved without special optimization for our model. Therefore, we want to check how well it can be improved by our ML based classifier, even if the parameter region is again excluded by $3l$–signal which is more sensitive. The $2l$–signal can be used to detect DM phenomenology, only if $2l$–signal without DM coupling is comparable with $3l$–signal according to our previous research [110]. In our previous studies, the addition of DM particles with charge–1 and charge–2 will not strongly enlarge the $2l$–signal comparing to the absence of DM case; hence there is no DM phenomenology from $2l$–signal, although it contains DM final states.

Similar to 3μ–signal we considered in previous section, here we discuss 2μ–signal, instead of $2l$–signal, as the simpler background. The old bound from $2l$–signal in our previous publication are again from 2μ–signal. Therefore, we are able to compare the new sensitivity with the old bounds directly. The result of 2μ–signal is shown in Fig. 4.6. The bound of 2μ–signal is weaker than 3μ–signal optimized by ML based classifier in all parameter region, but they are comparable. Therefore, if the coupling with DM is large enough, the sensitivity of 2μ–signal could be higher than the sensitivity of 3μ–signal. In contrast, if we assume the coupling with DM is 0, and hence the invisible decay channels of Z' are only $\bar{\nu}_\mu \nu_\mu$ and $\bar{\nu}_\tau \nu_\tau$, the whole parameter space related to 2μ final state has already been excluded by 3μ–signal similar to cut based analyses. Nevertheless, for $m_{Z'} < 100$ GeV, the best sensitivity is from the preselection $\not{E}_T > 10$ GeV, while for $m_{Z'} > 100$ GeV, the best sensitivity is from the preselection $\not{E}_T > 100$ GeV, and the sensitivity from 2μ–signal with ML based classifier is

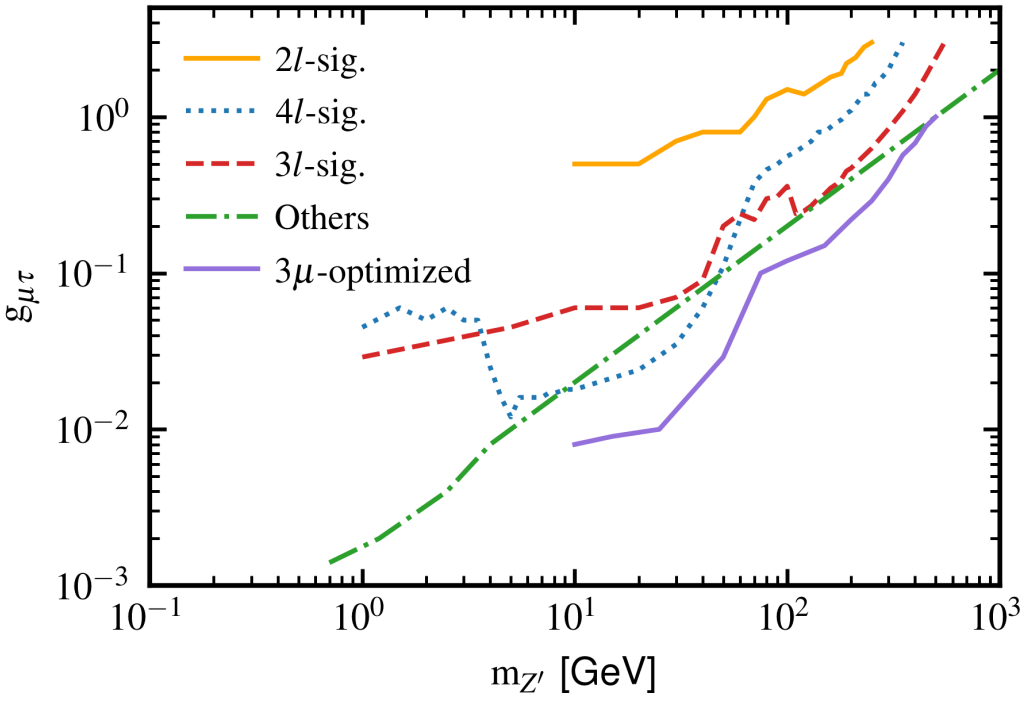

Figure 4.2: The best sensitivity for 3μ–signal in all ranges. For $m_{Z'} < 100$ GeV, the best results are from the preselection $\not{E}_T > 10$ GeV, while for $m_{Z'} > 100$ GeV, the best results are from the pre-selection $\not{E}_T > 100$ GeV. Moreover, we add data for $m_{Z'} = 15$, 25, 75, 150, 250, 350, and 450 GeV in validation set to ensure that our classifier is efficient for the untrained $m_{Z'}$.

better than the bound from the combination of all published analyses from ATLAS and CMS including $2l$, $3l$, and $4l$ final states.

To probe the DM phenomenology, we change the $L_\mu - L_\tau$ charge of DM from 0 to 1, 2, and 4, and check whether the classifier works well. The comparison is showed in Fig. 4.7 and Table 4.2. In Fig. 4.7, the increasing $L_\mu - L_\tau$ charges of DM lead the decrease of the number of events inside the peak, since more Z' decay invisibly, instead of $Z' \to \bar{l}l$. From the figure and more detailed information in the table, it is clear to see that our classifier can precisely select the signal data and distinguish different charges of DM. The only thing that may conceal the differences between various $L_\mu - L_\tau$ charges of DM is that the change of cross section is too small, and hence after the rescale, the differences are hard to probe.

In general, the $2l$–signal has the same behavior compared to 3μ–signal. The classification with pre–selection $\not{E}_T > 10$ GeV is better than the case with pre–selection $\not{E}_T > 100$ GeV only for $m_{Z'} < 100$ GeV. The reason is quite straightforward, which is exactly the same with the case in 3μ–signal, i.e. the data distributions in the phase space of $m_{Z'} < 100$ GeV and $m_{Z'} > 100$ GeV are different, while the preselection $\not{E}_T > 100$ GeV removes most of the events for $m_{Z'} < 100$ GeV and hence focuses on the heavy mediators. Although the missing momentum or the mass of the muon pair directly connect to $m_{Z'}$, they are not necessarily the most important features in Fig. 4.11, as the use of

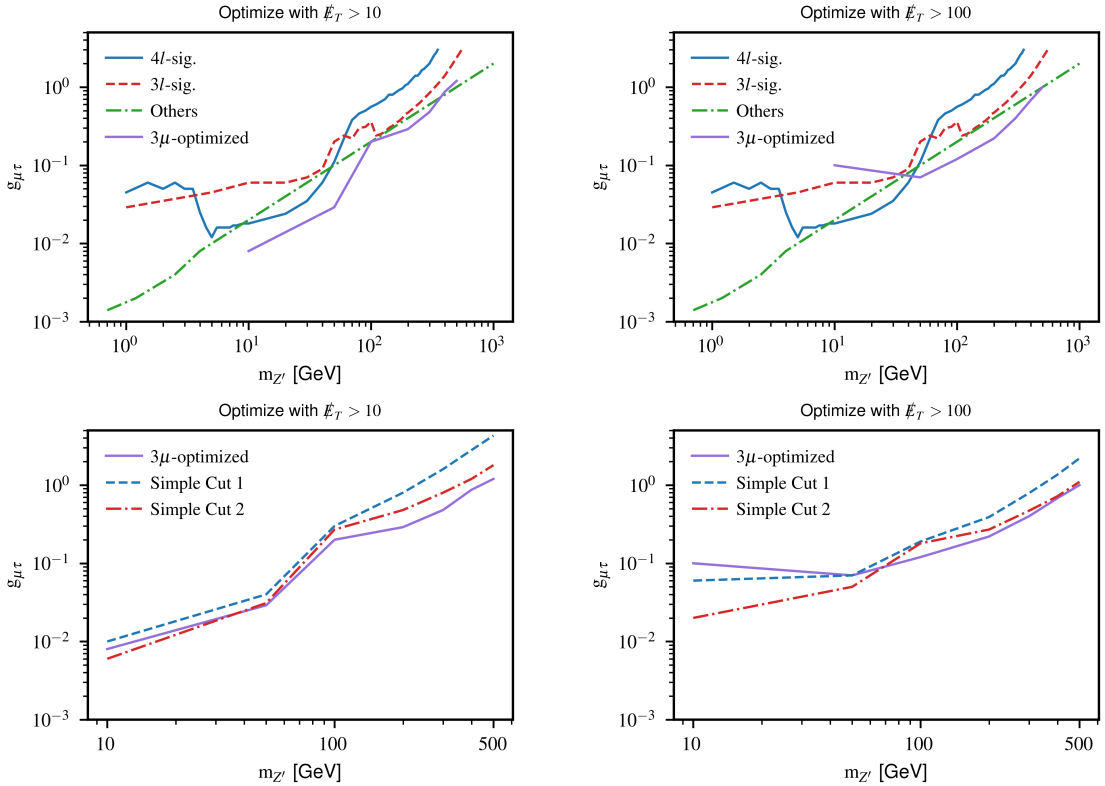

Figure 4.3: Improved bounds with pre–selections, where $\not{E}_T > 10$ GeV (left) and $\not{E}_T > 100$ GeV (right) respectively. The upper figures show the results from our previous publication (red, blue, green) [110] and ML classifiers (purple). In the lower figures, the optimized result through ML based classifier is compared to the results from 2 different dedicated cut schemes for every single mass point of $m_{Z'}$. The first one (SC1) only applies $|m^{(1)}_{\mu^+\mu^-} - m_Z| > 8$ GeV. The second one (SC2) applies an additional cut that $0.9\, m_{Z'} < m^{(Z')}_{\mu^+\mu^-} < 1.1\, m_{Z'}$. The cut varies for different $m_{Z'}$, and thus it outperforms the universal ML classifier at some special points.

Charge	180 GeV $< m_{\mu^+\mu^-} <$ 220 GeV	$m_{\mu^+\mu^-} <$ 180 GeV or $m_{\mu^+\mu^-} >$ 220 GeV
0	39615 (4543)	33493 (3841)
1	37937 (4477)	34691 (4094)
2	33110 (4220)	36648 (4671)
4	20392 (3147)	36184 (5584)

Table 4.2: The number of events selected by our preselection and classifier with (w/o) rescale according to cross sections in and out of the peak around $m_{Z'}$.

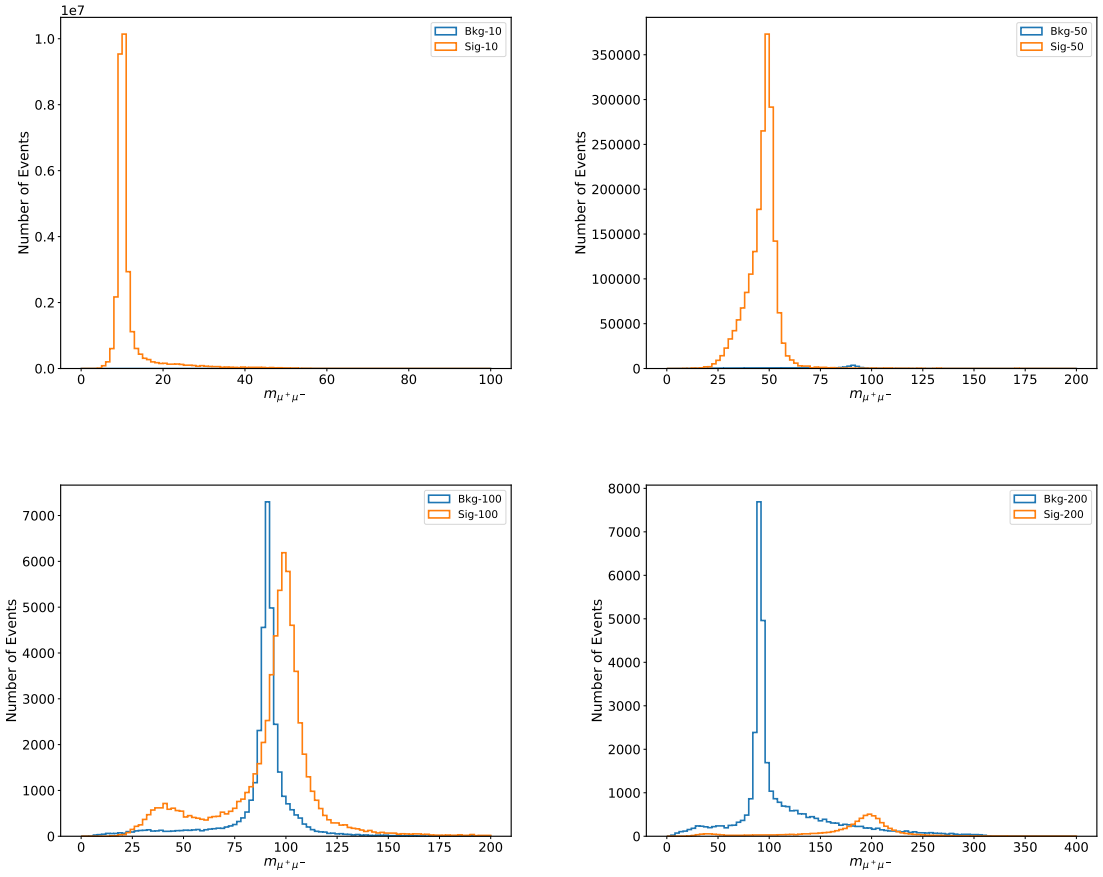

Figure 4.4: The figure show the distribution of $m^{(Z')}_{\mu^+\mu^-}$, which means the muon pair that has the invariant mass nearest to $m_{Z'}$. The y–axis for background is the event number after selection, while the y–axis of signal is the case after normalization according to the rate of cross section between background and signal to ensure the luminosities for both curves are the same. The top left is for $m_{Z'} = 10$ GeV, the top right is for $m_{Z'} = 50$ GeV, the bottom left is for $m_{Z'} = 100$ GeV, and the bottom right is for $m_{Z'} = 200$ GeV.

the same classifier for variant values of $m_{Z'}$ from 10 GeV to 500 GeV. Especially for $\not{E}_T$ in XGBoost, it is not one of the core features. On the other hand, we move to the case with preselection $\not{E}_T > 100$ GeV, $m_{\mu^+\mu^-}$ turns to be the most important feature in Fig. 4.12. In other words, the benefits of well grouped data help the training process. Nevertheless, the feature used in pre–selection may not works the best when distinguishing signal and backgrounds.

In summary, the pre–selection improves the data for training process (group similar data in phase space), while the features with high importance benefits the classification. In different parameter region, the better combination of those improves the final results. Only if the large enough data set with very complex neural network for Deep Learning are applied, the pre–selection for good enough classification of various parameter regions can be omitted by raw data with only 4 momenta. In contrast, our task is to use the ML algorithms to connect theoretical models with physical features, to achieve sufficient sensitivity for constraining the couplings, and to understand the topology of the event

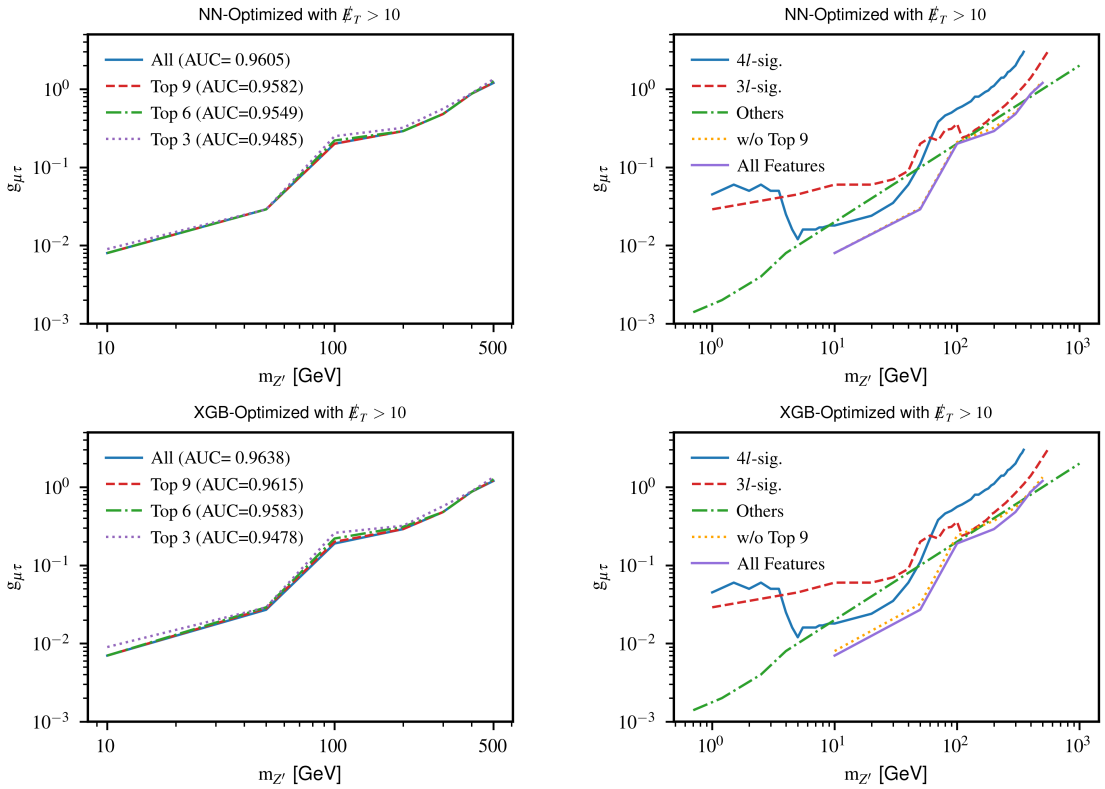

Figure 4.5: The figures show the new bound after removing part of the features used in previous ML models. The left one only keeps top 9/6/3 features in figure 4.8, while the right one removes the top 9 features. The upper figures use NN network, while the lower use XGBoost.

for designing collider searches. Therefore, the pre–selection for the separation of parameter regions, and the design of the features are still important. The combination of ML based algorithms and cut based event selection is widely used by ATLAS and CMS in their recent publications. Unfortunately, such analyses are hard to recast to study selected model for phenomenological researches. Our approach shows a new way that one can easily understand a theoretical model in detectors and estimate the sensitivity of collider data.

4.4 Conclusion

In this chapter, we use ML based classification to study simplified $U(1)_{L_\mu - L_\tau}$ extension of SM, containing a new mediator and DM particle. The simplified model is a simplification of a complete model that contains a gauge boson, a DM particle, an extra Higgs that offers mass for the new gauge boson and DM particle, and right handed neutrino with see–saw mechanism to generate tiny masses for left handed neutrinos. The analyses from ATLAS and CMS publications containing $2l$, $3l$, and $4l$ final states, where l means μ or hadronically decaying τ, can only probe the parameters in the simplified model. Although we have recast most of relevant LHC analyses, the bound is not good enough; hence we design ML based classifiers to distinguish signal and SM background to estimate

the sensitivity of LHC experiments.

Our classifier is outperformed in all parameter region compared to the previous result from both LHC and non–LHC analyses. In our ML based method, we mix XGBoost and Neural Network. On one hand, NN may helps us reach better results through a pure black box by the construction of all non–linear combination of the input vector; on the other hand, XGBoost tells us feature importance. We compare the results from both NN and XGBoost using only top 3/6/9 features in the feature importance from XGBoost. The general coincidence with classifiers trained by all features implies that the feature importance is physical, which corresponding to the properties of the theoretical model, rather than pure data tricks in algorithms. Moreover, although the high–level features are important in XGBoost classifier, they can be worked out internally by NN, as long as we keep low–level features that can construct high–level features, prepare large enough data set, and proceed well designed pre–selection. Therefore, the combination of tree based algorithms, which offers feature importance, and neural network, which usually gives the best classification, is a reasonable selection to read physical information of the theoretical model.

Through the distribution of top features according to importance, we found that the ML based classifier is flexible than cut based event selection, since the classifier can distinguish the difference of shapes between signal and background, while cut based method requires that the distribution should be centralized at different positions and not strongly overlapped. Moreover, different features are not necessarily used individually. The uses of different features are mixed together especially for fully connected neural network. Therefore, even if the signal and background are relatively overlapped, the ML algorithms can still offer good enough estimation of the distribution of signal.

The comparison between different pre–selections are discussed. In general, strong enough pre–selection for preventing outliers and removing irrelevant data is necessary. For both 3μ–signal and 2μ–signal, pre–selection $\not{E}_T > 10$ GeV is better for $m_{Z'} < 100$ GeV, while pre–selection $\not{E}_T > 100$ GeV is better for $m_{Z'} > 100$ GeV, given that $\not{E}_T > 100$ GeV removes most of events for $m_{Z'} < 100$ GeV and makes the classifier focus on the data with $m_{Z'} > 100$ GeV.

By removing part of the features from all candidates, we show that the information from top features can be reproduced by Deep Learning algorithms from low–level features like 4 momenta, if the data set is large enough. Nevertheless, only if reading the feature importance from tree based algorithms, we are able to know which features actually work in such a black box. Additionally, most of insensitive features can be removed without strongly influencing the result in both NN and XGBoost classifiers. Thus, when the result from NN is similar with the result from XGBoost, it means that we have already input the most important features, and are able to pick up the most significant parts through the feature importance. In this way, we can always get a set of core features that can offer good enough event selection and reflects the physical properties of a certain model. In cut based analysis, the physical information of the theoretical model is readable through the dedicated cuts designed and selected by experts. However, through ML based algorithms, the significant features that describe the topology of the new signal can be achieved by algorithms automatically. Therefore, it is an attractive way to understand theoretical model through data driven method.

Finally, the method can be extended to all final states that have already been researched by experimentalists to discovery the phenomenology of new models. It at least offers 2 kinds of information that is useful for phenomenological research. First is the minimal set of features that is able to distinguish the signal and background efficiently enough. Second is the estimated bound that can be achieved through the LHC data with selected final states in publications. It is a brief way to quickly access the phenomenology of any general simplified models by repeating the procedure on

selected models.

4.5 Figures for More Details

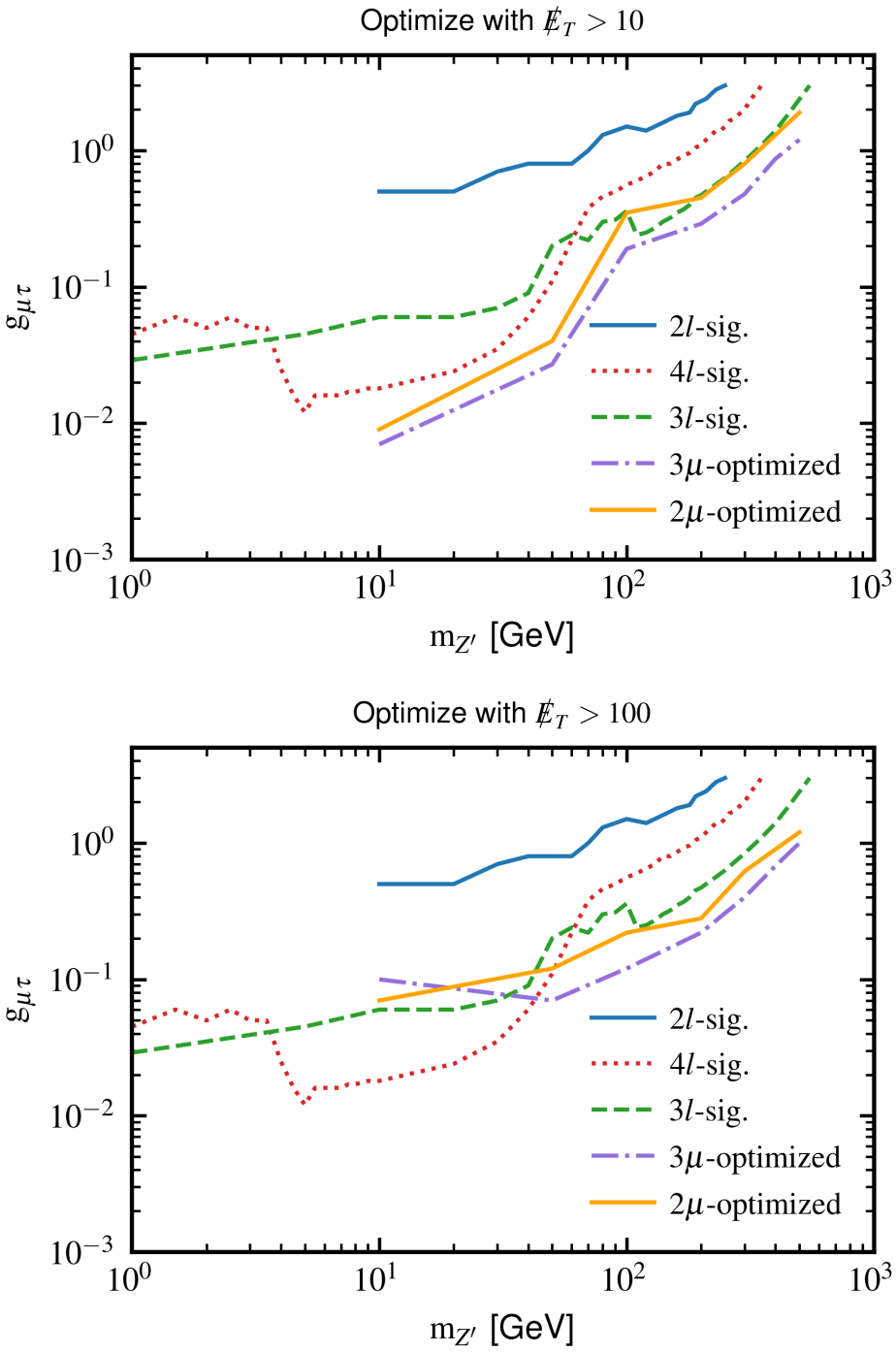

Figure 4.6: Improved sensitivity from $2l$–signal with pre–selections $\not{E}_T > 10$ GeV (upper) and $\not{E}_T > 100$ GeV (lower) respectively.

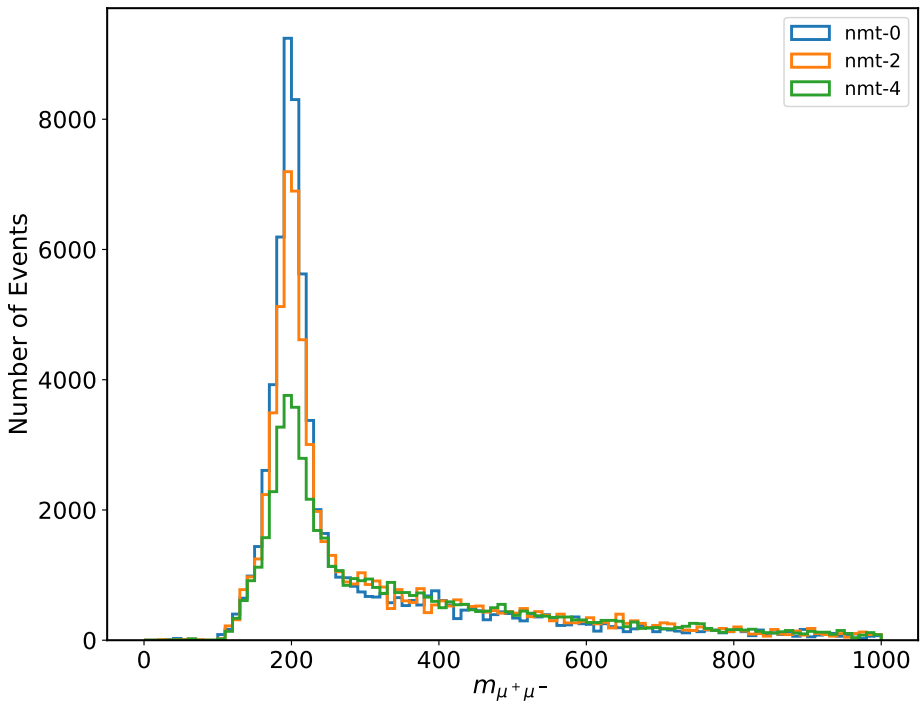

Figure 4.7: Distributions of $m_{\mu^+\mu^-}$ for 2μ–signal with $m_{Z'} = 200$ GeV and the DM carrying various $L_\mu - L_\tau$ charges equaling to 0, 2, and 4. The y–axis is rescaled according to the cross sections to keep the same luminosities for 3 histograms.

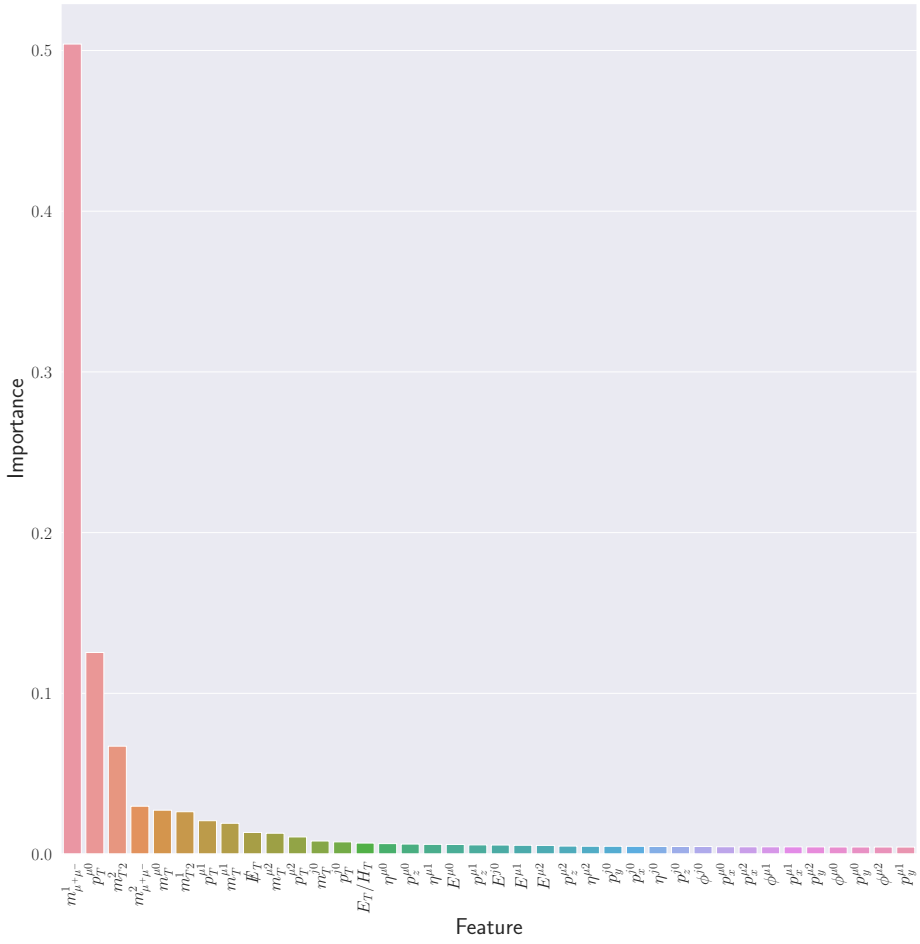

Figure 4.8: Feature importance for $3l$–signal with pre–selection $E\!\!\!/_T > 10$ GeV from XGBoost, where the parameter is described in table 4.1. The most influential features are $m_{\mu^+\mu^-}$, $p_T^{\mu_0}$, $p_T^{\mu_1}$, $m_T^{\mu_0}$, $m_T^{\mu_1}$, $m_{T2}^{\mu^+\mu^-}$, $E\!\!\!/_T$, $m_T^{\mu_2}$, and $p_T^{\mu_2}$.

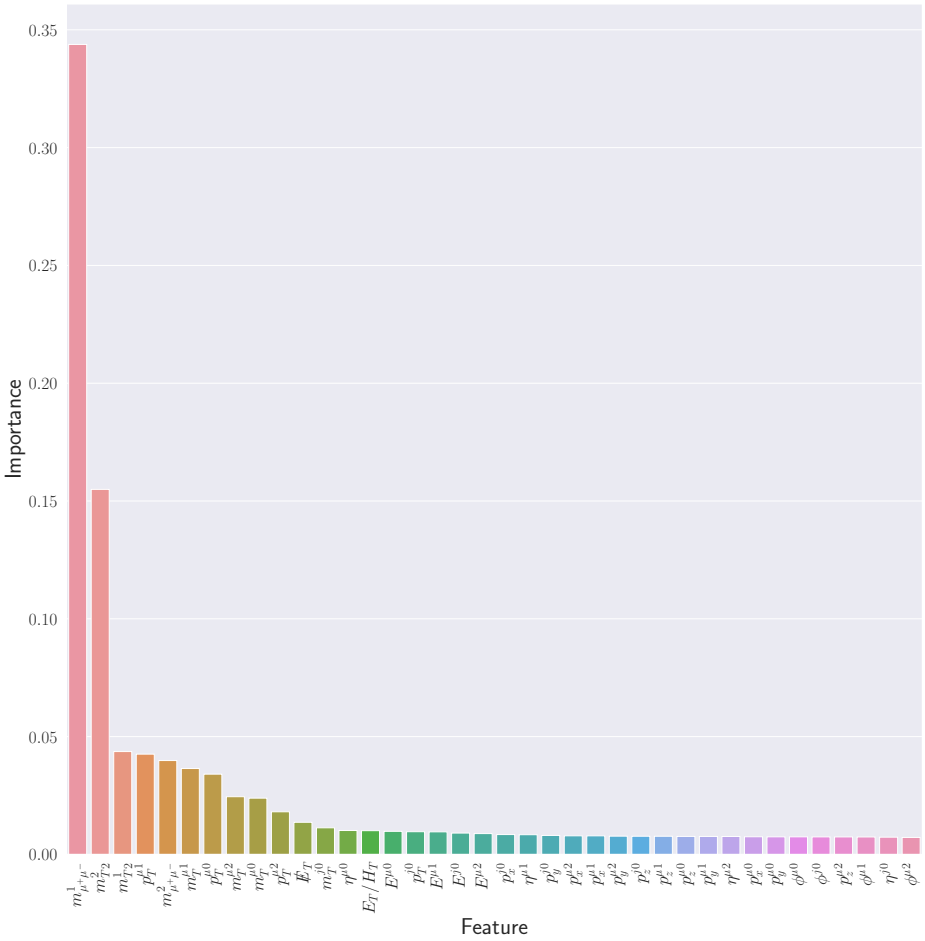

Figure 4.9: Feature importance for $3l$–signal with pre–selection $E\!\!\!/_T > 100$ GeV from XGBoost, where the parameter is described in table 4.1. The most influential features are $m^{(1)}_{\mu^+\mu^-}$, $m^{(2)}_{T2}$, $m^{(1)}_{T2}$, $p^{\mu_1}_T$, $m^{(2)}_{\mu^+\mu^-}$, $m^{\mu_1}_T$, $p^{\mu_0}_T$, $m^{\mu_2}_T$, and $m^{\mu_0}_T$.

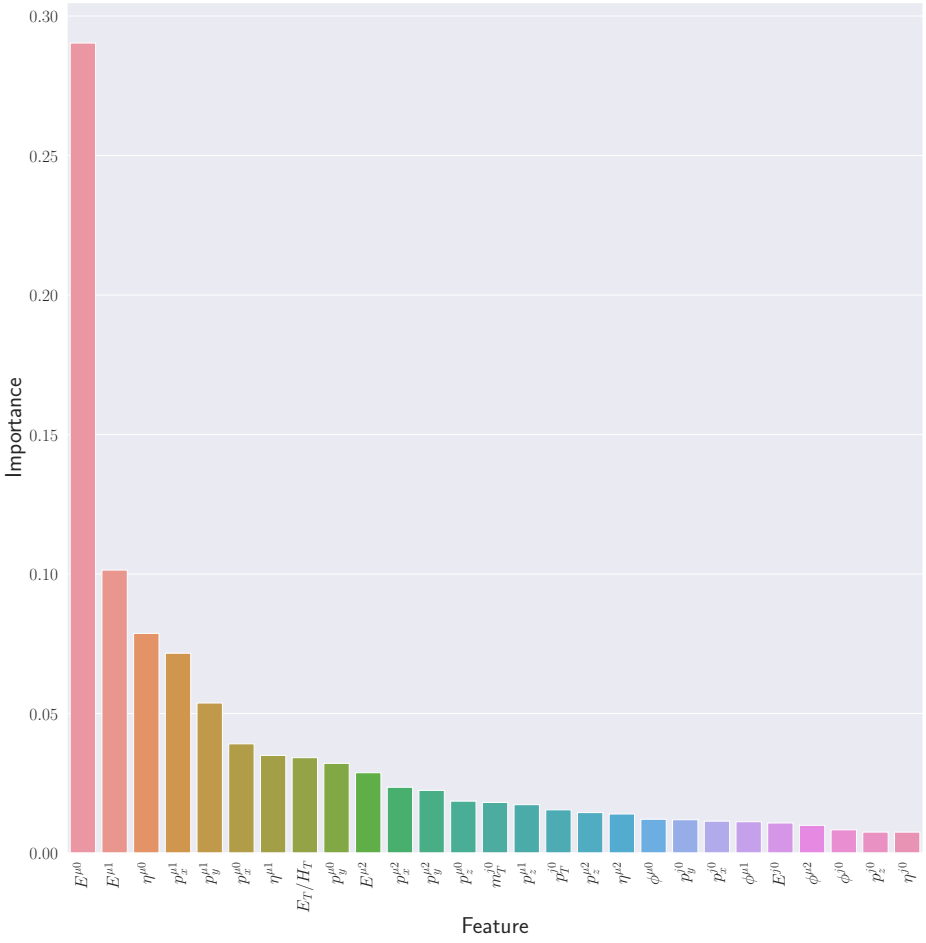

Figure 4.10: Feature importance for $3l$–signal with pre–selection $E_T^{miss} > 10$ GeV and exclusion of the top 9 features in XGBoost, where the parameters are described in table 4.1. The most influential features are E^{μ_0}, E^{μ_1}, η^{μ_0}, $p_x^{\mu_1}$, $p_y^{\mu_1}$, $p_x^{\mu_0}$, η^{μ_1}, E_T/H_T, and $p_y^{\mu_0}$.

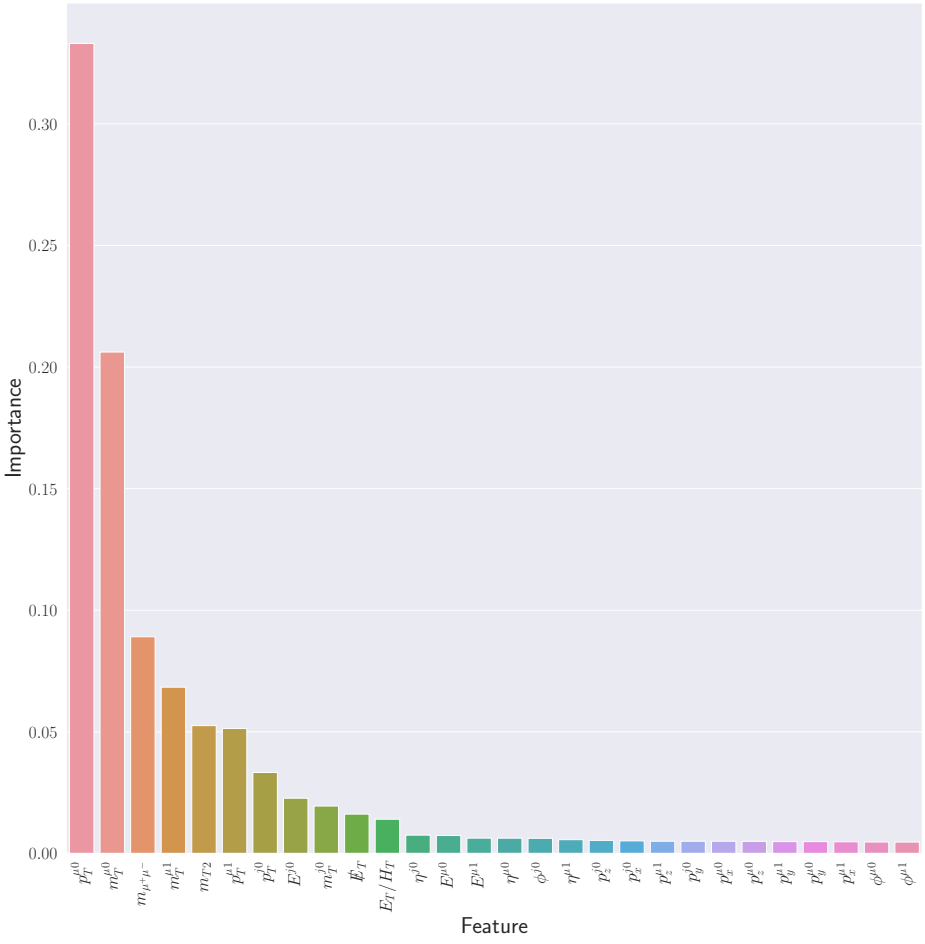

Figure 4.11: Feature importance for $2l$–signal with pre–selection $E\!\!\!/_T > 10$ GeV from XGBoost, where the parameter is described in table 4.1. The most influential features are $p_T^{\mu_0}$, $m_T^{\mu_0}$, $m_{\mu^+\mu^-}$, $m_T^{\mu_1}$, m_{T2}, $p_T^{\mu_1}$, $p_T^{j_0}$, E^{j_0}, $m_T^{j_0}$, and $E\!\!\!/_T$.

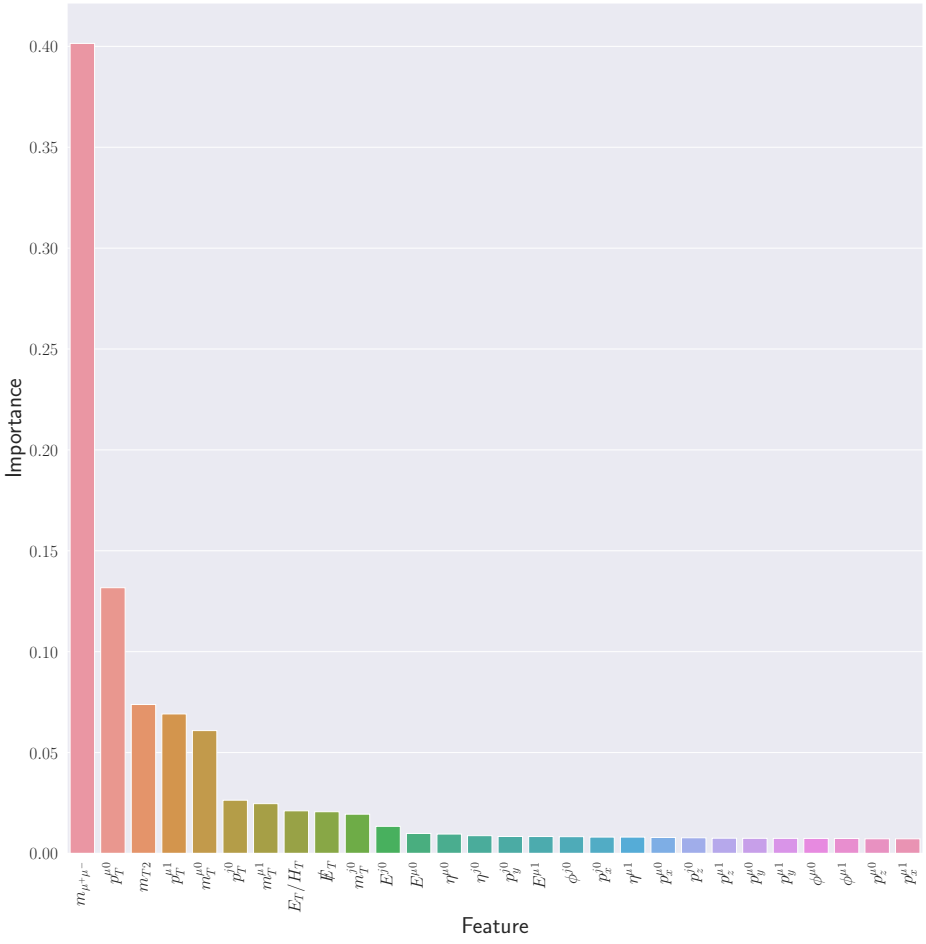

Figure 4.12: Feature importance for $2l$–signal with pre–selection $E\!\!\!/_T > 100$ GeV from XGBoost, where the parameter is described in table 4.1. The most influential features are $m_{\mu^+\mu^-}$, $p_T^{\mu_0}$, m_{T2}, $p_T^{\mu_1}$, $m_T^{\mu_0}$, $p_T^{j_0}$, $m_T^{\mu_1}$, E_T/H_T, $E\!\!\!/_T$ and $m_T^{j_0}$.

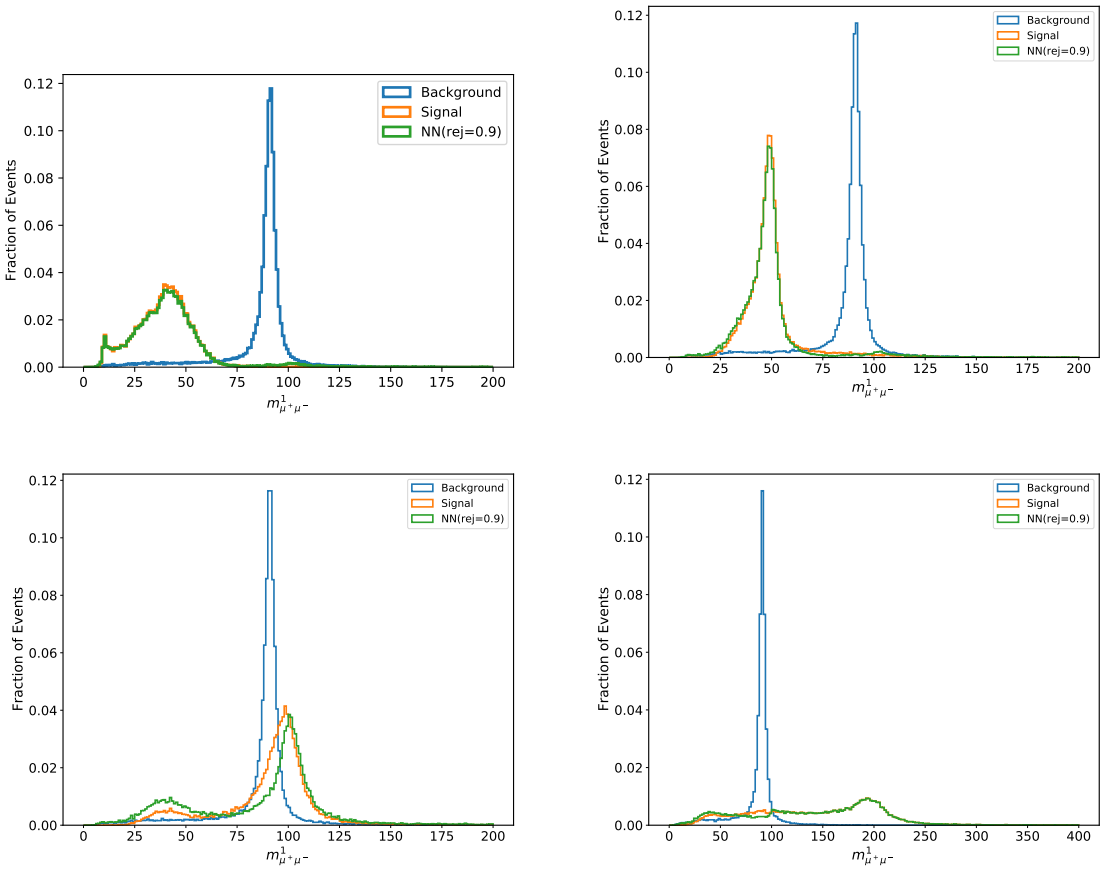

Figure 4.13: The histograms of signal and background for 3μ final state with different $m_{Z'}$ according to $m^{(1)}_{\mu^+\mu^-}$ bins. In the upper left feature, $m_{Z'} = 10$ GeV. In the upper right figure, $m_{Z'} = 50$ GeV. In the lower left figure, $m_{Z'} = 100$ GeV. In the lower right figure, $m_{Z'} = 200$ GeV.

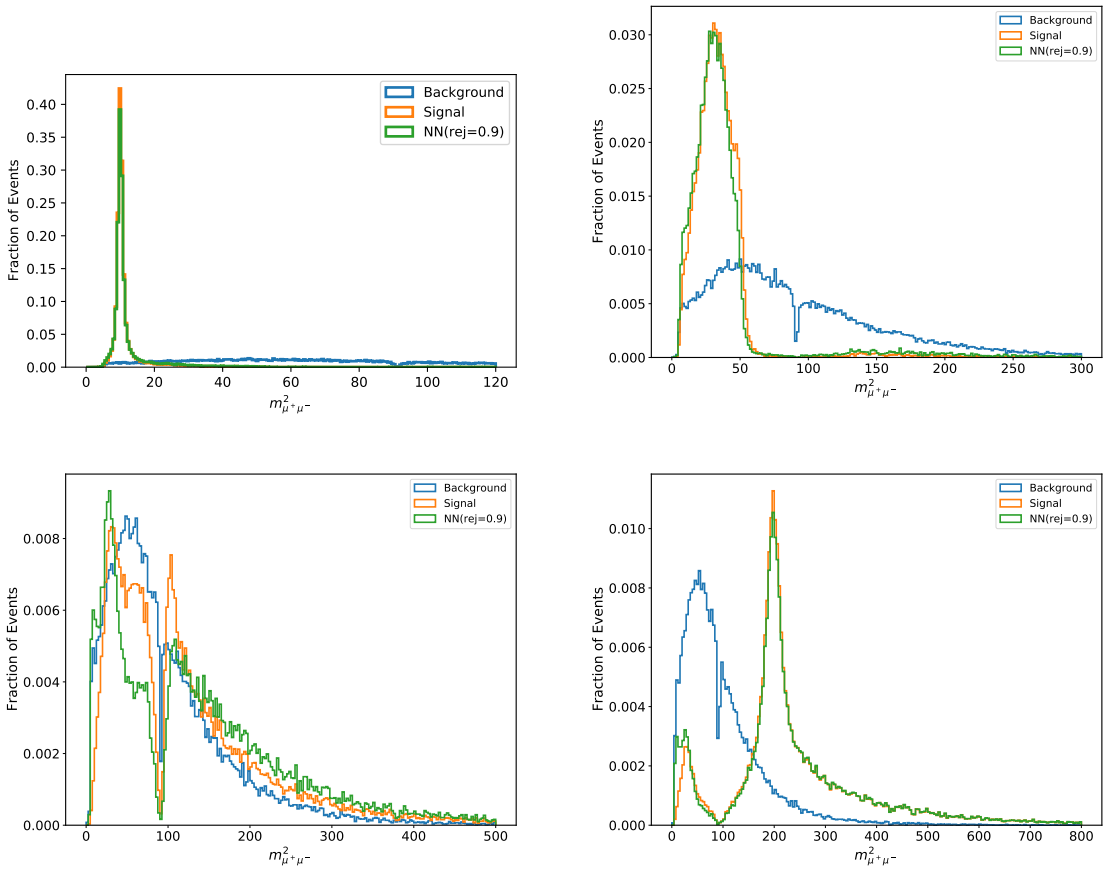

Figure 4.14: The histograms of signal and background for 3μ final state with different $m_{Z'}$ according to $m^{(2)}_{\mu^+\mu^-}$ bins. In the upper left feature, $m_{Z'} = 10$ GeV. In the upper right figure, $m_{Z'} = 50$ GeV. In the lower left figure, $m_{Z'} = 100$ GeV. In the lower right figure, $m_{Z'} = 200$ GeV.

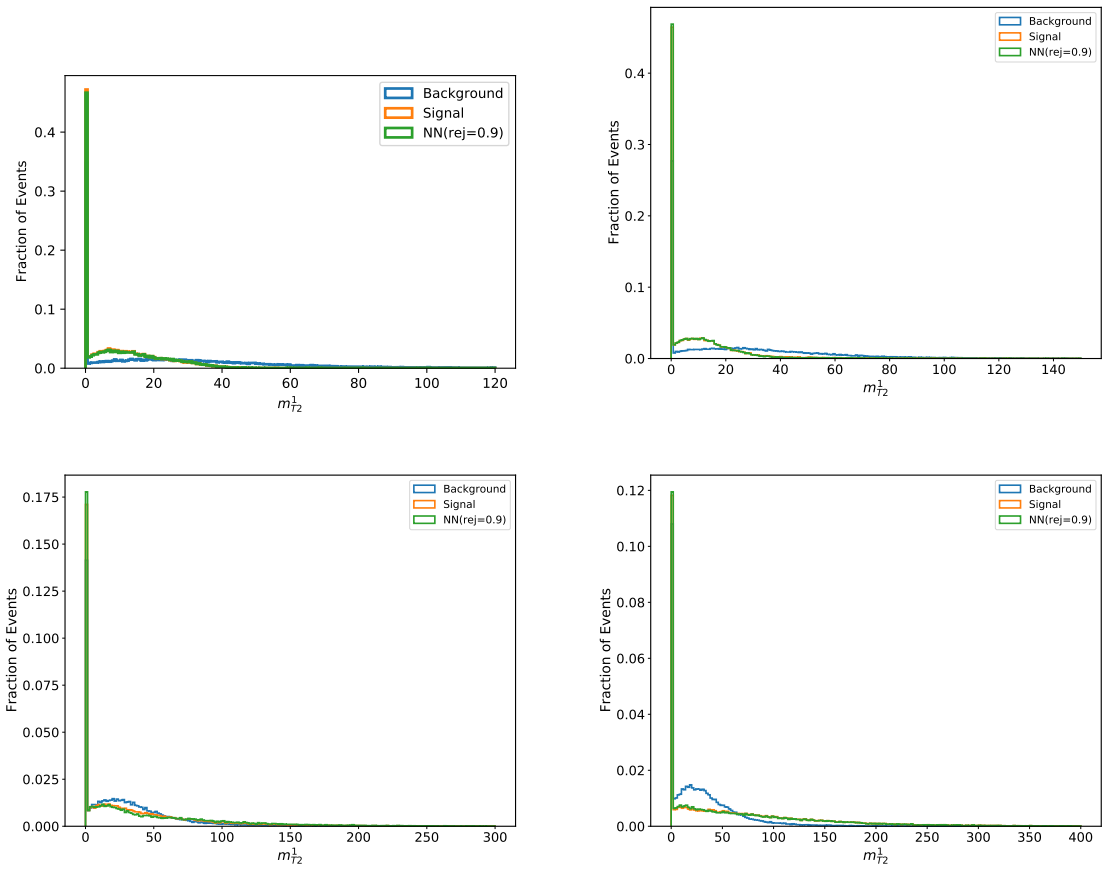

Figure 4.15: The histograms of signal and background for 3μ final state with different $m_{Z'}$ according to $m_{T2}^{(1)}$ bins. In the upper left feature, $m_{Z'} = 10$ GeV. In the upper right figure, $m_{Z'} = 50$ GeV. In the lower left figure, $m_{Z'} = 100$ GeV. In the lower right figure, $m_{Z'} = 200$ GeV.

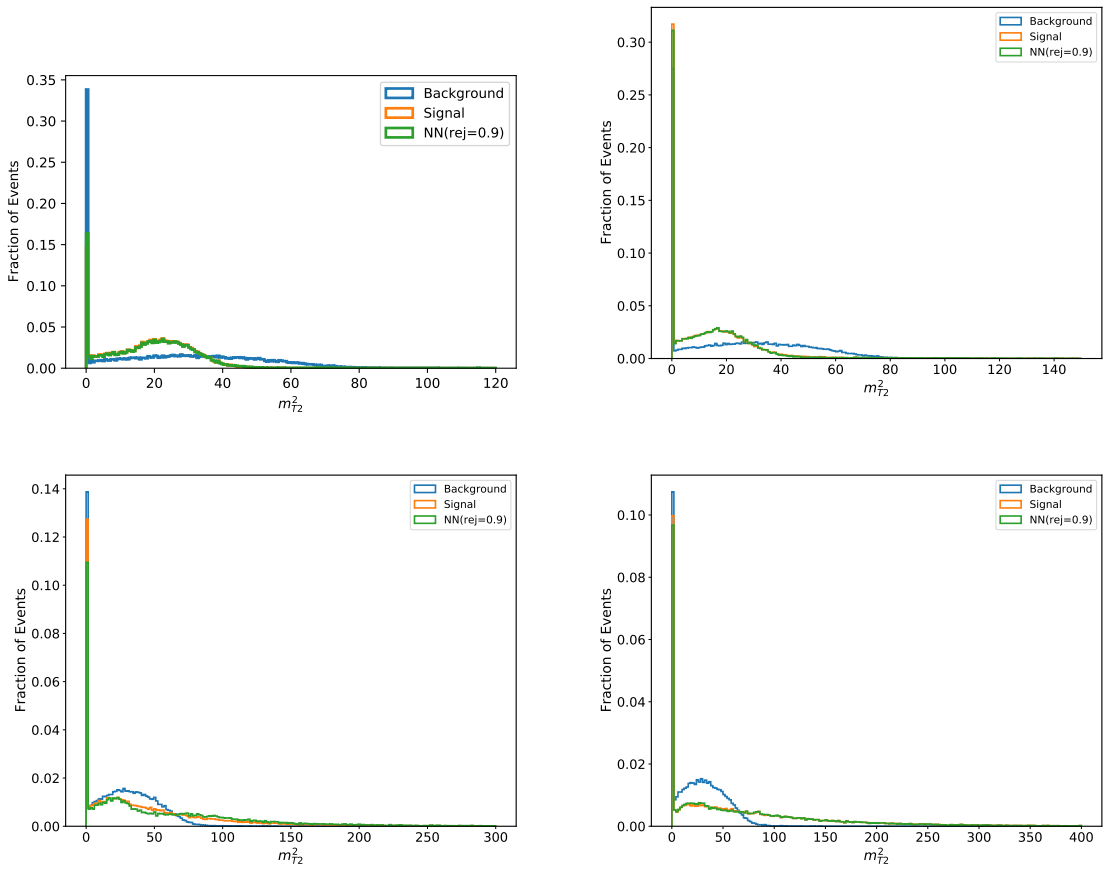

Figure 4.16: The histograms of signal and background for 3μ final state with different $m_{Z'}$ according to $m^{(2)}_{T2}$ bins. In the upper left feature, $m_{Z'} = 10$ GeV. In the upper right figure, $m_{Z'} = 50$ GeV. In the lower left figure, $m_{Z'} = 100$ GeV. In the lower right figure, $m_{Z'} = 200$ GeV.

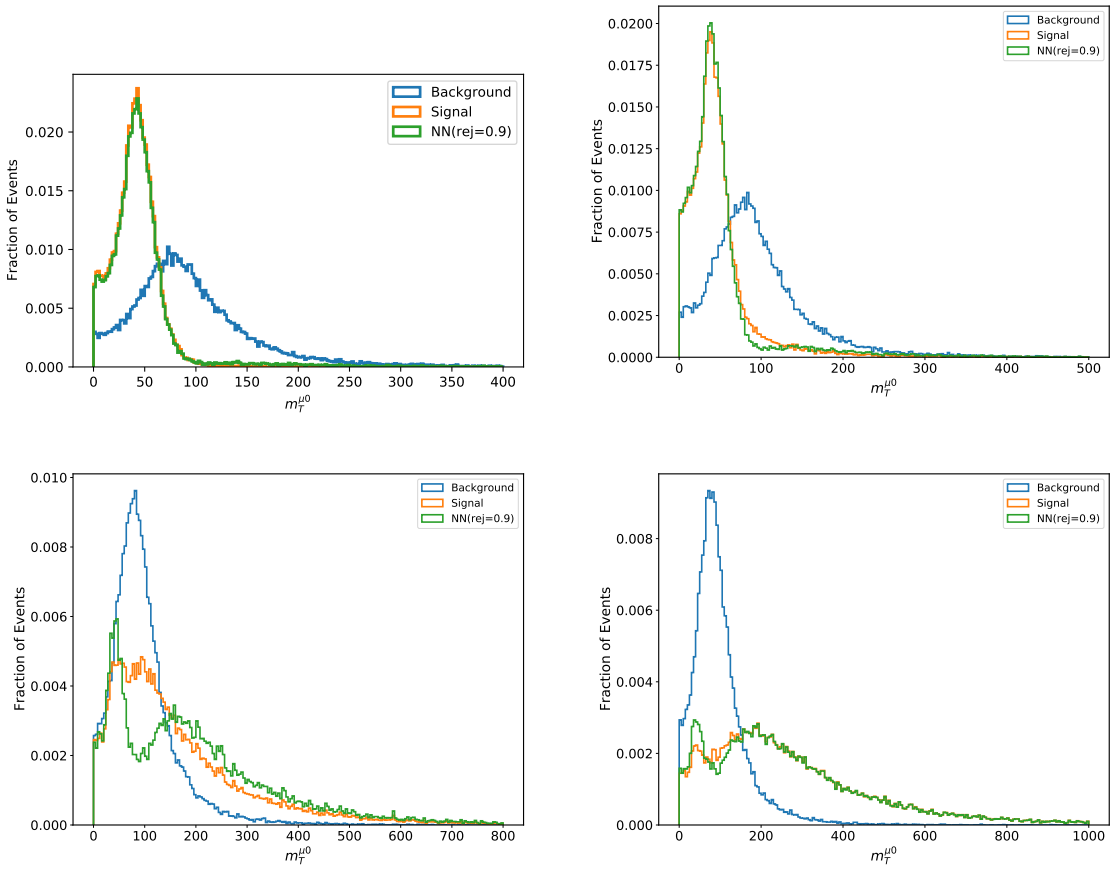

Figure 4.17: The histograms of signal and background for 3μ final state with different $m_{Z'}$ according to m_T^0 bins. In the upper left feature, $m_{Z'} = 10$ GeV. In the upper right figure, $m_{Z'} = 50$ GeV. In the lower left figure, $m_{Z'} = 100$ GeV. In the lower right figure, $m_{Z'} = 200$ GeV.

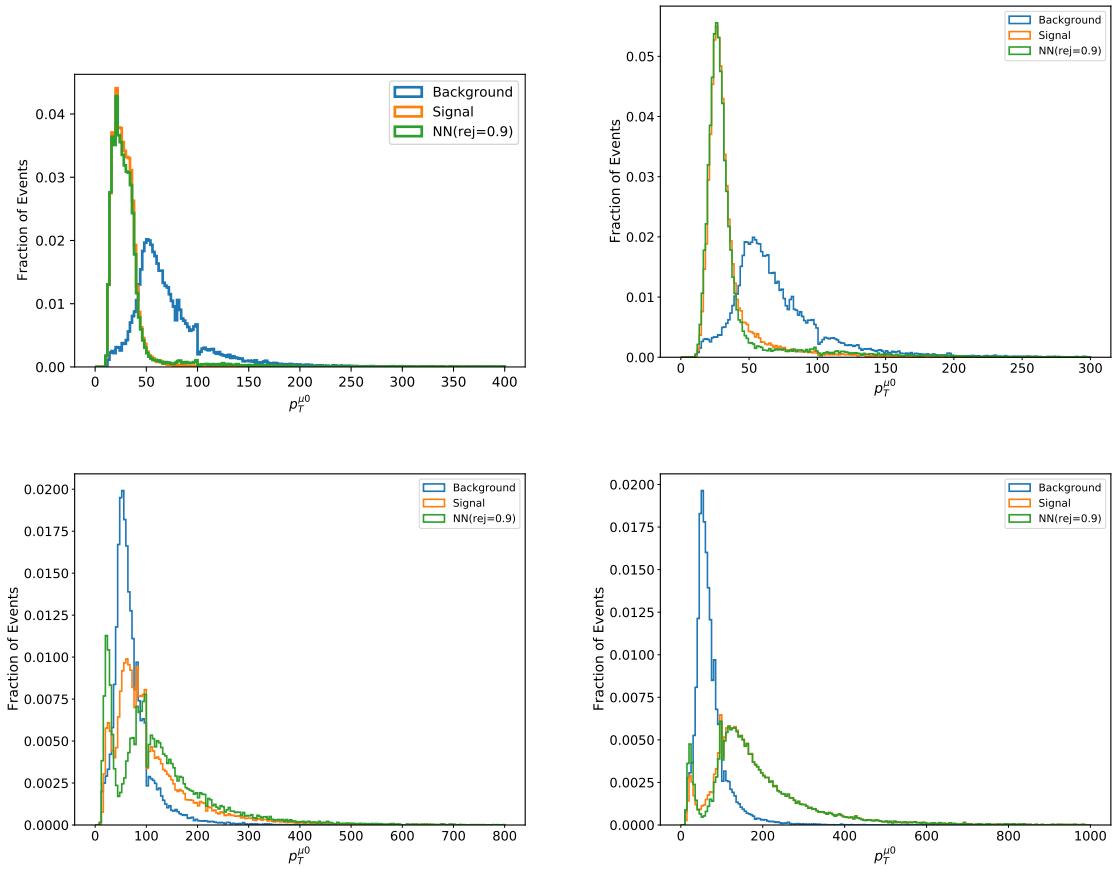

Figure 4.18: The histograms of signal and background for 3μ final state with different $m_{Z'}$ according to p_T^0 bins. In the upper left feature, $m_{Z'} = 10$ GeV. In the upper right figure, $m_{Z'} = 50$ GeV. In the lower left figure, $m_{Z'} = 100$ GeV. In the lower right figure, $m_{Z'} = 200$ GeV.

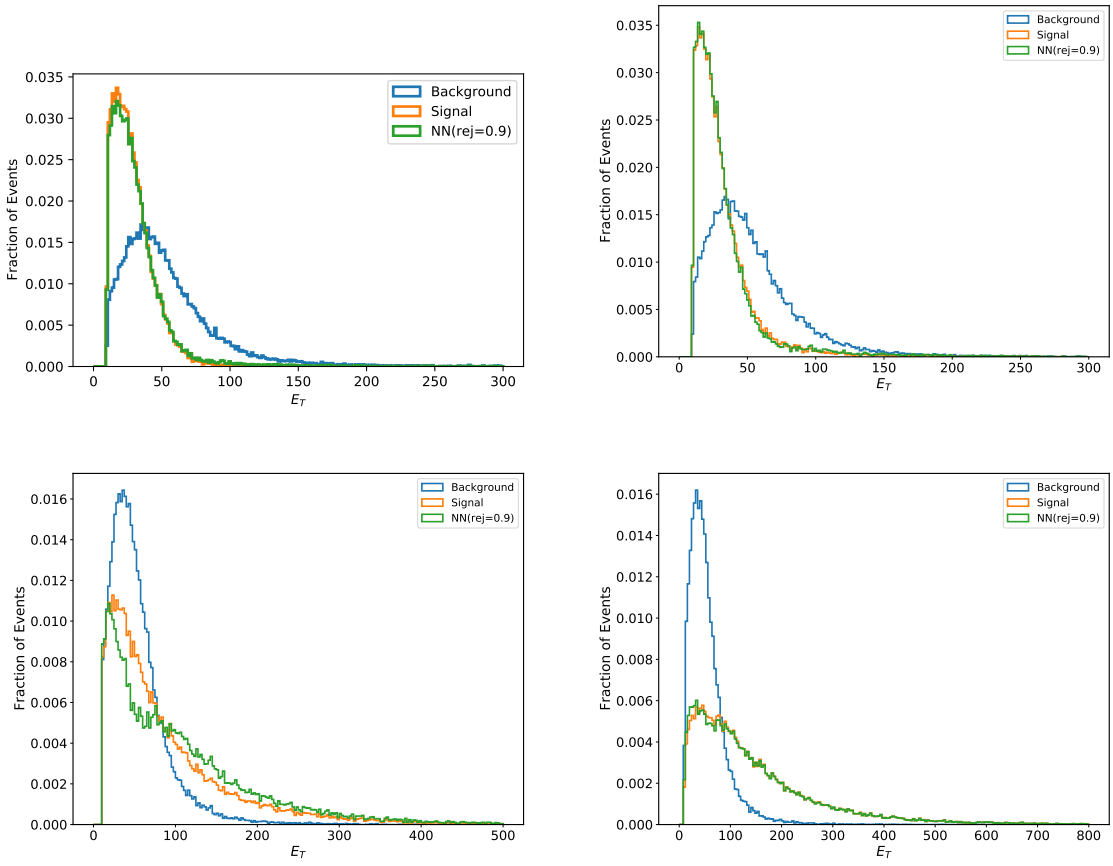

Figure 4.19: The histograms of signal and background for 3μ final state with different $m_{Z'}$ according to $\not{E}_T$ bins. In the upper left feature, $m_{Z'} = 10$ GeV. In the upper right figure, $m_{Z'} = 50$ GeV. In the lower left figure, $m_{Z'} = 100$ GeV. In the lower right figure, $m_{Z'} = 200$ GeV.

CHAPTER 5

Conclusions and Outlook

$U(1)_{L_\mu - L_\tau}$ model extends the Standard Model (SM) by a new gauge group without introducing gauge anomalies; the lack of extra couplings to electrons allows it escaping from searches like Large Electron-Positron Collider. Since muon is charged under the new gauge group, it also contributes sizable corrections to the anomalies magnetic moment of the muon, $g_\mu - 2$, which is larger than the SM prediction by about 3.5σ. Besides, without loss of generality, an extra complex scalar field ϕ_{DM} can be incorporated into this model as a viable dark matter (DM) candidate, whose nature is still unknown for us. The experimental searches for this model include final states with multi-leptons plus invisible particles (such as neutrinos and DM), which makes Large Hadron Collider (LHC) a good choice to study its constraints. Even though there is a CMS analysis [65] specially designed for searching the model, it only covers new Z' masses between 5 and 70 GeV based on 4μ final states, and most of the LHC data still remain unexplored. In this thesis, we studied the constraints on $U(1)_{L_\mu - L_\tau}$ model from LHC data by using `CheckMATE` framework and Machine Learning (ML) based methods.

In Chapter 1 we briefly talked about the ingredients of the SM, which has made a great success in the last decades. We also discussed its limitations, such as the lack of DM candidate, hierarchy problem, incomparable with gravity, etc. To test the SM and search for signals beyond it, we introduced the LHC as a good resource since its high energy level (up to 14 TeV) and high integrated luminosity (up to 139 fb^{-1}). Then we showed the standard procedures to test this model against LHC data, as well as the possibility to using a ML based method.

In Chapter 2 we reviewed the history and discovery on the existence of DM. It started with the mismatch on the rotation curve of spiral galaxies, and was almost conclusive due to the findings on Bullet Cluster. Then we focused on the Weakly Interacting Massive Particles (WIMPs) as DM candidates, and recalled their constraints in different experimental searches, including direct detection, indirect detection and collider searches.

In Chapter 3 we first discussed different kinds of $U(1)$ gauge extensions including $B - L$, $L_\mu - L_\tau$, $L_e - L_\mu$ and $L_e - L_\tau$. Except $L_\mu - L_\tau$, the others are highly constrained by searches in LEP. The main ingredients of this model include a new gauge boson Z' and a complex scalar ϕ_{DM} for DM particle, two Higgs-like scalars, and three extra fermions for neutrino mixing. We neglected the interference between Z' and Z/γ, as well as the mixing in Higgs sectors. Since the right-handed neutrinos are too heavy to produce at LHC, they are also not in our consideration. Then the remaining free parameters are the Z' mass $m_{Z'}$ and the new coupling $g_{\mu\tau}$, which locate in the parameter region respecting to the perturbative condition. Based on this formalism, we studied the constraints on the

simplified $L_\mu - L_\tau$ model using the CheckMATE framework; in total, we encoded 281 new signal regions defined in 28 different LHC analyses papers. For comparison, we also derived upper bounds from low energy non-LHC experiments (such as the CCFR and BaBar) using the CL_S method. The final state signals we considered for LHC analyses include $2l$, $3l$, or $4l$ plus missing energy (neutrinos and DM particles), where l means a muon or hadronically decaying τ. We found that, on the one hand, $2l$ final states are always less sensitive than $3l$ and $4l$ final states, regardless that DM particles are included or not, which means that LHC are not sensitive to the production of DM particles in this model. On the other hand, the non-LHC bounds are generally stronger than LHC bounds, except for 10 GeV $< m_{Z'} <$ 60 GeV. In this region, the best bound comes from a dedicated CMS analysis specially designed to search for $L_\mu - L_\tau$ gauge boson, which is based on 4μ final states. However, outside the range of 10 GeV $< m_{Z'} <$ 60 GeV, the final LHC limit is set by $3l$ final states. This implies that by optimizing the cuts, it is likely to have stronger LHC constraints, especially for 3μ final states.

In Chapter 4 we focused on optimizing the cuts of $U(1)_{L_\mu - L_\tau}$ model with final states of 3μ and 2μ plus DM particles. We neglected 4μ final state since it was shown less sensitive than 3μ state in our previous study. The distinguish between signal and background events is performed by ML classifiers: Gradient Boosting Decision Tree (GBDT) and deep learning neural network (NN); the inputs are physical variables (like four-momenta and di-muon invariant mass) extracted by CheckMATE; the final bound is derived by null hypothesis test. It is important to emphasize that, for each kind of classifier, we only train a single classifier over the whole range of Z' masses. For comparison, we also use a dedicated cut for different Z' masses separately as references. Then we applied pre-selection and prepared data as two sets: $\not{E}_T > 10$ GeV and $\not{E}_T > 100$ GeV; this is necessary for improving the performance and reducing outliers in samples. We found that, ML classifiers outperform in all parameter region compared to previous results from both LHC and non-LHC analyses; they also perform better in most of $m_{Z'}$ points than our dedicated cuts; for $m_{Z'} < 100$ GeV, classifier trained on $\not{E}_T > 10$ GeV works better, while for $m_{Z'} > 100$ GeV, $\not{E}_T > 100$ GeV is better; moreover, they perform almost equally well at Z' mass which never appears in training set. On the other hand, to uncover the "black-box" of ML methods, we ranked the feature importance extracted from GBDT method, and only selected top $3/6/9$ features as inputs. However, the sensitivity only decreases a little even for the top 3 case; this indicates that, the feature importance indeed reflects physical information of the distribution in collider, not just a pure data tricks in algorithm. To further address the physical insights of the ML classifiers, we compare the normalized event distribution of simulated signal, background and model predicted signal at different features; it turns out that the predicted signal and background are indeed well separated.

LHC has provided us a large number of valuable and unexplored data for exploring and testing new physics. In this thesis, we use $U(1)_{L_\mu - L_\tau}$ model as a template to test model against the LHC data; the methods applied in this thesis can be easily generalized to other models. We show that model test can be done either by traditional cut-based "recast", or by ML-based methods. Rather than just being a "dark magic", the feature importance extracted from ML method indeed has its physical insight; and it could be helpful while designing and selecting physical variables during recast.

Bibliography

[1] T. Aoyama et al., *Tenth-Order Electron Anomalous Magnetic Moment — Contribution of Diagrams without Closed Lepton Loops*,
Phys. Rev. D **91** (2015) 033006, [Erratum: Phys.Rev.D 96, 019901 (2017)],
arXiv: 1412.8284 [hep-ph] (cit. on p. 1).

[2] D. Hanneke, S. Fogwell and G. Gabrielse,
New Measurement of the Electron Magnetic Moment and the Fine Structure Constant,
Phys. Rev. Lett. **100** (12 2008) 120801,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.100.120801 (cit. on p. 1).

[3] M. D. Schwartz, *Quantum Field Theory and the Standard Model*,
Cambridge University Press, 2014, ISBN: 978-1-107-03473-0, 978-1-107-03473-0
(cit. on p. 1).

[4] Y. Grossman, "Introduction to flavor physics",
2009 European School of High-Energy Physics, 2010 111, arXiv: 1006.3534 [hep-ph]
(cit. on p. 1).

[5] J. Zupan, *Introduction to flavour physics*,
CERN Yellow Rep. School Proc. **6** (2019) 181, ed. by M. Mulders and C. Duhr,
arXiv: 1903.05062 [hep-ph] (cit. on p. 1).

[6] S. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579 (cit. on p. 1).

[7] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367
(cit. on p. 1).

[8] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264 (cit. on p. 1).

[9] G. Aad et al., *Combined search for the Standard Model Higgs boson using up to 4.9 fb^{-1} of pp collision data at $\sqrt{s}$ = 7 TeV with the ATLAS detector at the LHC*,
Phys. Lett. B **710** (2012) 49, arXiv: 1202.1408 [hep-ex] (cit. on pp. 1, 4, 5).

[10] S. Chatrchyan et al., *Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s}$ = 7 TeV*, Phys. Lett. B **710** (2012) 26, arXiv: 1202.1488 [hep-ex]
(cit. on pp. 1, 4, 5).

[11] J. Woithe, G. J. Wiener and F. F. Van der Veken,
Let's have a coffee with the Standard Model of particle physics!,
Phys. Educ. **52** (2017) 034001 (cit. on p. 2).

[12] S. Cunliffe, *Flavor Physics*, 2020,
URL: https://indico.desy.de/event/26403/contributions/60314/ (cit. on p. 3).

[13] A. Riotto and M. Trodden, *Recent progress in baryogenesis*, Ann. Rev. Nucl. Part. Sci. **49** (1999) 35, arXiv: hep-ph/9901362 (cit. on p. 3).

[14] R. Aaij et al., *Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$*, JHEP **09** (2015) 179, arXiv: 1506.08777 [hep-ex] (cit. on p. 3).

[15] R. Aaij et al., *Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays*, Phys. Rev. Lett. **122** (2019) 191801, arXiv: 1903.09252 [hep-ex] (cit. on p. 3).

[16] H. P. Nilles, *Supersymmetry, Supergravity and Particle Physics*, Phys. Rept. **110** (1984) 1 (cit. on p. 4).

[17] M. Drees, R. Godbole and P. Roy, *Theory and phenomenology of sparticles: An account of four-dimensional N=1 supersymmetry in high energy physics*, 2004 (cit. on p. 4).

[18] P. de Salas et al., *Status of neutrino oscillations 2018: 3σ hint for normal mass ordering and improved CP sensitivity*, Phys. Lett. B **782** (2018) 633, arXiv: 1708.01186 [hep-ph] (cit. on p. 4).

[19] R. N. Mohapatra and G. Senjanovic, *Neutrino Mass and Spontaneous Parity Nonconservation*, Phys. Rev. Lett. **44** (1980) 912 (cit. on p. 4).

[20] M. J. Dolinski, A. W. Poon and W. Rodejohann, *Neutrinoless Double-Beta Decay: Status and Prospects*, Ann. Rev. Nucl. Part. Sci. **69** (2019) 219, arXiv: 1902.04097 [nucl-ex] (cit. on p. 4).

[21] N. Jarosik et al., *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results*, Astrophys. J. Suppl. **192** (2011) 14, arXiv: 1001.4744 [astro-ph.CO] (cit. on p. 4).

[22] K. Garrett and G. Duda, *Dark Matter: A Primer*, Adv. Astron. **2011** (2011) 968283, arXiv: 1006.2483 [hep-ph] (cit. on pp. 4, 11).

[23] M. Drees et al., *CheckMATE: Confronting your Favourite New Physics Model with LHC Data*, Comput. Phys. Commun. **187** (2015) 227, arXiv: 1312.2591 [hep-ph] (cit. on pp. 7, 36).

[24] D. Dercks et al., *CheckMATE 2: From the model to the limit*, Comput. Phys. Commun. **221** (2017) 383, arXiv: 1611.09856 [hep-ph] (cit. on pp. 7, 26, 36, 37).

[25] J. Alwall et al., *MadGraph 5: going beyond*, Journal of High Energy Physics **2011** (2011) 1 (cit. on pp. 7, 26, 36).

[26] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications **191** (2015) 159 (cit. on pp. 7, 27, 36).

[27] J. de Favereau et al., *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, JHEP **02** (2014) 057, arXiv: 1307.6346 [hep-ex] (cit. on pp. 7, 28, 36).

[28] A. L. Read, *Presentation of search results: The CL(s) technique*, J. Phys. **G28** (2002) 2693, [,11(2002)] (cit. on pp. 7, 36).

[29] J. Schmidhuber, *Deep Learning in Neural Networks: An Overview*, arXiv e-prints, arXiv:1404.7828 (2014) arXiv:1404.7828, arXiv: 1404.7828 [cs.NE] (cit. on p. 7).

[30] T. Young et al., *Recent Trends in Deep Learning Based Natural Language Processing*, arXiv e-prints, arXiv:1708.02709 (2017) arXiv:1708.02709, arXiv: 1708.02709 [cs.CL] (cit. on p. 7).

[31] E. Aprile et al., *Physics reach of the XENON1T dark matter experiment*, JCAP **04** (2016) 027, arXiv: 1512.07501 [physics.ins-det] (cit. on p. 9).

[32] E. Aprile et al., *XENON1T dark matter data analysis: Signal and background models and statistical inference*, Phys. Rev. D **99** (2019) 112009, arXiv: 1902.11297 [physics.ins-det] (cit. on p. 9).

[33] D. Hooper and L. Goodenough, *Dark Matter Annihilation in The Galactic Center As Seen by the Fermi Gamma Ray Space Telescope*, Phys. Lett. B **697** (2011) 412, arXiv: 1010.2752 [hep-ph] (cit. on p. 9).

[34] G. Bertone and D. Hooper, *History of dark matter*, Rev. Mod. Phys. **90** (4 2018) 045002, URL: https://link.aps.org/doi/10.1103/RevModPhys.90.045002 (cit. on pp. 9, 10).

[35] V. C. Rubin and J. Ford W.Kent, *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, Astrophys. J. **159** (1970) 379 (cit. on p. 10).

[36] K. Begeman, *H I rotation curves of spiral galaxies. I - NGC 3198*, Astron. Astrophys. **223** (1989) 47 (cit. on pp. 10, 11).

[37] D. Clowe et al., *A direct empirical proof of the existence of dark matter*, Astrophys. J. Lett. **648** (2006) L109, arXiv: astro-ph/0608407 (cit. on p. 13).

[38] C. Bennett et al., *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results*, Astrophys. J. Suppl. **208** (2013) 20, arXiv: 1212.5225 [astro-ph.CO] (cit. on p. 14).

[39] N. Aghanim et al., *Planck 2018 results. VI. Cosmological parameters*, Astron. Astrophys. **641** (2020) A6, arXiv: 1807.06209 [astro-ph.CO] (cit. on p. 15).

[40] M. Schumann, *Direct Detection of WIMP Dark Matter: Concepts and Status*, J. Phys. G **46** (2019) 103003, arXiv: 1903.03026 [astro-ph.CO] (cit. on pp. 15–17).

[41] D. Akerib et al., *Results from a search for dark matter in the complete LUX exposure*, Phys. Rev. Lett. **118** (2017) 021303, arXiv: 1608.07648 [astro-ph.CO] (cit. on p. 16).

[42] X. Cui et al., *Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment*, Phys. Rev. Lett. **119** (2017) 181302, arXiv: 1708.06917 [astro-ph.CO] (cit. on p. 16).

[43] P. Agnes et al., *Low-Mass Dark Matter Search with the DarkSide-50 Experiment*, Phys. Rev. Lett. **121** (2018) 081307, arXiv: 1802.06994 [astro-ph.HE] (cit. on p. 16).

[44] R. Agnese et al., *First Dark Matter Constraints from a SuperCDMS Single-Charge Sensitive Detector*, Phys. Rev. Lett. **121** (2018) 051301, [Erratum: Phys.Rev.Lett. 122, 069901 (2019)], arXiv: 1804.10697 [hep-ex] (cit. on p. 16).

[45] E. Aprile et al., *Dark Matter Search Results from a One Ton-Year Exposure of XENON1T*, Phys. Rev. Lett. **121** (2018) 111302, arXiv: 1805.12562 [astro-ph.CO] (cit. on p. 16).

[46] J. M. Gaskins, *A review of indirect searches for particle dark matter*, Contemp. Phys. **57** (2016) 496, arXiv: 1604.00014 [astro-ph.HE] (cit. on p. 16).

[47] R. K. Leane, "Indirect Detection of Dark Matter in the Galaxy", *3rd World Summit on Exploring the Dark Side of the Universe*, 2020, arXiv: 2006.00513 [hep-ph] (cit. on pp. 16, 19).

[48] L. Goodenough and D. Hooper, *Possible Evidence For Dark Matter Annihilation In The Inner Milky Way From The Fermi Gamma Ray Space Telescope*, (2009), arXiv: 0910.2998 [hep-ph] (cit. on p. 18).

[49] V. Lefranc and E. Moulin, "Dark matter search in the inner galactic center halo with H.E.S.S", *51st Rencontres de Moriond on Cosmology*, 2016 149, arXiv: 1608.08453 [astro-ph.HE] (cit. on p. 19).

[50] H. Abdallah et al., *Search for γ-Ray Line Signals from Dark Matter Annihilations in the Inner Galactic Halo from 10 Years of Observations with H.E.S.S.*, Phys. Rev. Lett. **120** (2018) 201101, arXiv: 1805.05741 [astro-ph.HE] (cit. on p. 19).

[51] A. Boveia and C. Doglioni, *Dark Matter Searches at Colliders*, Ann. Rev. Nucl. Part. Sci. **68** (2018) 429, arXiv: 1810.12238 [hep-ex] (cit. on p. 18).

[52] L. Basso, S. Moretti and G. M. Pruna, *Phenomenology of the minimal $B-L$ extension of the Standard Model: the Higgs sector*, Phys. Rev. D **83** (2011) 055014, arXiv: 1011.2612 [hep-ph] (cit. on pp. 22, 23).

[53] A. M. Sirunyan et al., *Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, Phys. Rev. Lett. **122** (2019) 121803, arXiv: 1811.09689 [hep-ex] (cit. on p. 23).

[54] T. Robens, T. Stefaniak and J. Wittbrodt, *Two-real-scalar-singlet extension of the SM: LHC phenomenology and benchmark scenarios*, Eur. Phys. J. C **80** (2020) 151, arXiv: 1908.08554 [hep-ph] (cit. on p. 23).

[55] T. Stefaniak, "Higgs physics beyond the Standard Model", *An Alpine LHC Physics Summit 2019*, 2019, arXiv: 1908.10900 [hep-ph] (cit. on p. 23).

[56] S. Amrith et al., *LHC Constraints on a $B-L$ Gauge Model using Contur*, JHEP **05** (2019) 154, arXiv: 1811.11452 [hep-ph] (cit. on p. 23).

[57] A. Biswas, S. Choubey and S. Khan, *Neutrino Mass, Dark Matter and Anomalous Magnetic Moment of Muon in a $U(1)_{L_\mu - L_\tau}$ Model*, JHEP **09** (2016) 147, arXiv: 1608.04194 [hep-ph] (cit. on p. 24).

[58] M. Carena et al., *Z' gauge bosons at the Tevatron*, Phys. Rev. D **70** (2004) 093009, arXiv: hep-ph/0408098 (cit. on p. 24).

[59] J. Heeck, *Unbroken $B-L$ symmetry*, Phys. Lett. B **739** (2014) 256, arXiv: 1408.6845 [hep-ph] (cit. on p. 24).

[60] K. Asai et al., *Minimal Gauged $U(1)_{L_\alpha - L_\beta}$ Models Driven into a Corner*, Phys. Rev. D **99** (2019) 055029, arXiv: 1811.07571 [hep-ph] (cit. on p. 24).

[61] M. Bauer et al., *Dark Matter in Anomaly-Free Gauge Extensions*, SciPost Phys. **5** (2018) 036, arXiv: 1805.01904 [hep-ph] (cit. on p. 24).

[62] A. Biswas, S. Choubey and S. Khan, *Neutrino Mass, Dark Matter and Anomalous Magnetic Moment of Muon in a $U(1)_{L_\mu - L_\tau}$ Model*, JHEP **09** (2016) 147, arXiv: 1608.04194 [hep-ph] (cit. on pp. 24, 33).

[63] A. Biswas, S. Choubey and S. Khan, *FIMP and Muon $(g-2)$ in a $U(1)_{L_\mu - L_\tau}$ Model*, JHEP **02** (2017) 123, arXiv: 1612.03067 [hep-ph] (cit. on p. 24).

[64] J. L. Rainbolt and M. Schmitt, *Branching fraction for Z decays to four leptons and constraints on new physics*, (2018), arXiv: 1805.05791 [hep-ex] (cit. on p. 25).

[65] A. M. Sirunyan et al., *Search for an $L_\mu - L_\tau$ gauge boson using $Z \to 4\mu$ events in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Submitted to: Phys. Lett. (2018), arXiv: 1808.03684 [hep-ex] (cit. on pp. 25, 27–29, 32, 33, 35, 63).

[66] F. del Aguila et al., *Collider limits on leptophilic interactions*, JHEP **03** (2015) 059, arXiv: 1411.7394 [hep-ph] (cit. on p. 25).

[67] K. Harigaya et al., *Muon g-2 and LHC phenomenology in the $L_\mu - L_\tau$ gauge symmetric model*, JHEP **03** (2014) 105, arXiv: 1311.0870 [hep-ph] (cit. on p. 25).

[68] F. Elahi and A. Martin, *Constraints on $L_\mu - L_\tau$ interactions at the LHC and beyond*, Phys. Rev. D **93** (1 2016) 015022, URL: https://link.aps.org/doi/10.1103/PhysRevD.93.015022 (cit. on p. 25).

[69] A. Alloul et al., *FeynRules 2.0 - A complete toolbox for tree-level phenomenology*, Comput. Phys. Commun. **185** (2014) 2250, arXiv: 1310.1921 [hep-ph] (cit. on p. 26).

[70] V. Khachatryan et al., *Search for a standard model-like Higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the LHC*, Phys. Lett. **B744** (2015) 184, arXiv: 1410.6679 [hep-ex] (cit. on p. 27).

[71] S. Chatrchyan et al., *Measurement of the W^+W^- Cross section in pp Collisions at $\sqrt{s} = 7$ TeV and Limits on Anomalous $WW\gamma$ and WWZ couplings*, Eur. Phys. J. **C73** (2013) 2610, arXiv: 1306.1126 [hep-ex] (cit. on p. 27).

[72] V. Khachatryan et al., *Measurement of the transverse momentum spectrum of the Higgs boson produced in pp collisions at $\sqrt{s} = 8$ TeV using $H \to WW$ decays*, JHEP **03** (2017) 032, arXiv: 1606.01522 [hep-ex] (cit. on p. 27).

[73] V. Khachatryan et al., *Measurement of the $\mathrm{W^+W^-}$ cross section in pp collisions at $\sqrt{s} = 8$ TeV and limits on anomalous gauge couplings*, Eur. Phys. J. **C76** (2016) 401, arXiv: 1507.03268 [hep-ex] (cit. on p. 27).

[74] G. Aad et al., *Search for type-III Seesaw heavy leptons in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector*, Phys. Rev. **D92** (2015) 032001, arXiv: 1506.01839 [hep-ex] (cit. on p. 27).

[75] A. M. Sirunyan et al., *Search for dark matter and unparticles in events with a Z boson and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **03** (2017) 061, [Erratum: JHEP09,106(2017)], arXiv: 1701.02042 [hep-ex] (cit. on p. 27).

[76] A. M. Sirunyan et al., *Search for new phenomena in final states with two opposite-charge, same-flavor leptons, jets, and missing transverse momentum in pp collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **03** (2018) 076, arXiv: 1709.08908 [hep-ex] (cit. on p. 27).

[77] A. M. Sirunyan et al., *Search for new physics in events with a leptonically decaying Z boson and a large transverse momentum imbalance in proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. **C78** (2018) 291, arXiv: 1711.00431 [hep-ex] (cit. on p. 27).

[78] A. M. Sirunyan et al., *Search for new physics in events with two soft oppositely charged leptons and missing transverse momentum in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, Phys. Lett. **B782** (2018) 440, arXiv: 1801.01846 [hep-ex] (cit. on pp. 27, 29).

[79] A. M. Sirunyan et al., *Combined search for electroweak production of charginos and neutralinos in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **03** (2018) 160, arXiv: 1801.03957 [hep-ex] (cit. on pp. 27, 29).

[80] A. M. Sirunyan et al., *Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at* $\sqrt{s} = 13$ *TeV*, Submitted to: Phys. Lett. (2018), arXiv: 1806.05246 [hep-ex] (cit. on pp. 27, 29).

[81] A. M. Sirunyan et al., *Search for high-mass resonances in dilepton final states in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **06** (2018) 120, arXiv: 1803.06292 [hep-ex] (cit. on p. 27).

[82] A. M. Sirunyan et al., *Search for supersymmetric partners of electrons and muons in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, Submitted to: Phys. Lett. (2018), arXiv: 1806.05264 [hep-ex] (cit. on p. 27).

[83] A. M. Sirunyan et al., *Searches for pair production of charginos and top squarks in final states with two oppositely charged leptons in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, Submitted to: JHEP (2018), arXiv: 1807.07799 [hep-ex] (cit. on pp. 27, 29).

[84] M. Aaboud et al., *Search for electroweak production of supersymmetric particles in final states with two or three leptons at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, (2018), arXiv: 1803.02762 [hep-ex] (cit. on p. 27).

[85] M. Aaboud et al., *Constraints on off-shell Higgs boson production and the Higgs boson total width in* $ZZ \to 4\ell$ *and* $ZZ \to 2\ell 2\nu$ *final states with the ATLAS detector*, (2018), arXiv: 1808.01191 [hep-ex] (cit. on p. 27).

[86] A. M. Sirunyan et al., *Search for dark matter produced in association with a Higgs boson decaying to* $\gamma\gamma$ *or* $\tau^+\tau^-$ *at* $\sqrt{s} = 13$ *TeV*, JHEP **09** (2018) 046, arXiv: 1806.04771 [hep-ex] (cit. on p. 27).

[87] A. M. Sirunyan et al., *Search for supersymmetry in events with a* τ *lepton pair and missing transverse momentum in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, (2018), arXiv: 1807.02048 [hep-ex] (cit. on p. 27).

[88] M. Aaboud et al., *Search for additional heavy neutral Higgs and gauge bosons in the ditau final state produced in 36* fb^{-1} *of pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **01** (2018) 055, arXiv: 1709.07242 [hep-ex] (cit. on p. 27).

[89] M. Aaboud et al., *Search for lepton-flavor violation in different-flavor, high-mass final states in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Submitted to: Phys. Rev. (2018), arXiv: 1807.06573 [hep-ex] (cit. on p. 27).

[90] V. Khachatryan et al., *A search for pair production of new light bosons decaying into muons*, Phys. Lett. **B752** (2016) 146, arXiv: 1506.00424 [hep-ex] (cit. on p. 27).

[91] A. M. Sirunyan et al., *Measurements of the* pp → ZZ *production cross section and the* Z → 4ℓ *branching fraction, and constraints on anomalous triple gauge couplings at* $\sqrt{s}$ = 13 *TeV*, Eur. Phys. J. **C78** (2018) 165, [Erratum: Eur. Phys. J.C78,no.6,515(2018)], arXiv: 1709.08601 [hep-ex] (cit. on pp. 27, 29).

[92] A. M. Sirunyan et al., *Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s}$ = 13 TeV*, JHEP **08** (2018) 066, arXiv: 1803.05485 [hep-ex] (cit. on p. 27).

[93] M. Aaboud et al., *$ZZ \to \ell^+\ell^-\ell'^+\ell'^-$ cross-section measurements and search for anomalous triple gauge couplings in 13 TeV pp collisions with the ATLAS detector*, Phys. Rev. **D97** (2018) 032005, arXiv: 1709.07703 [hep-ex] (cit. on p. 27).

[94] M. Aaboud et al., *Measurement of the Higgs boson coupling properties in the $H \to ZZ^* \to 4\ell$ decay channel at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **03** (2018) 095, arXiv: 1712.02304 [hep-ex] (cit. on p. 27).

[95] M. Aaboud et al., *Search for doubly charged scalar bosons decaying into same-sign W boson pairs with the ATLAS detector*, Submitted to: Eur. Phys. J. (2018), arXiv: 1808.01899 [hep-ex] (cit. on p. 27).

[96] A. M. Sirunyan et al., *Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state of two muons and two τ leptons in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, (2018), arXiv: 1805.04865 [hep-ex] (cit. on p. 27).

[97] A. M. Sirunyan et al., *Search for electroweak production of charginos and neutralinos in multilepton final states in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, JHEP **03** (2018) 166, arXiv: 1709.05406 [hep-ex] (cit. on pp. 27, 29).

[98] V. Khachatryan et al., *Search for light bosons in decays of the 125 GeV Higgs boson in proton-proton collisions at $\sqrt{s}$ = 8 TeV*, JHEP **10** (2017) 076, arXiv: 1701.02032 [hep-ex] (cit. on p. 27).

[99] F. Jegerlehner and A. Nyffeler, *The Muon g-2*, Phys. Rept. **477** (2009) 1, arXiv: 0902.3360 [hep-ph] (cit. on p. 28).

[100] S. R. Mishra et al., *Neutrino tridents and W Z interference*, Phys. Rev. Lett. **66** (1991) 3117 (cit. on pp. 28, 33).

[101] J. P. Lees et al., *Search for a muonic dark force at BABAR*, Phys. Rev. **D94** (2016) 011102, arXiv: 1606.03501 [hep-ex] (cit. on pp. 28, 33).

[102] W. Altmannshofer et al., *Neutrino Trident Production: A Powerful Probe of New Physics with Neutrino Beams*, Phys. Rev. Lett. **113** (2014) 091801, arXiv: 1406.2332 [hep-ph] (cit. on pp. 28, 33).

[103] D. Geiregat et al., *First observation of neutrino trident production*, Phys. Lett. **B245** (1990) 271 (cit. on pp. 28, 33).

[104] E. J. Chun et al., *Searching for flavored gauge bosons*, JHEP **02** (2019) 093, [Erratum: JHEP 07, 024 (2019)], arXiv: 1811.04320 [hep-ph] (cit. on p. 28).

[105] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250 (cit. on p. 28).

[106] X. G. He et al., *NEW Z-prime PHENOMENOLOGY*, Phys. Rev. **D43** (1991) 22 (cit. on p. 33).

[107] M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001 (cit. on p. 33).

[108] K. Asai, K. Hamaguchi and N. Nagata, *Predictions for the neutrino parameters in the minimal gauged $U(1)_{L_\mu - L_\tau}$ model*, Eur. Phys. J. C **77** (2017) 763, arXiv: 1705.00419 [hep-ph] (cit. on p. 33).

[109] K. Asai et al., *Minimal Gauged $U(1)_{L_\alpha - L_\beta}$ Models Driven into a Corner*, Phys. Rev. D **99** (2019) 055029, arXiv: 1811.07571 [hep-ph] (cit. on p. 33).

[110] M. Drees, M. Shi and Z. Zhang, *Constraints on $U(1)_{L_\mu - L_\tau}$ from LHC Data*, Phys. Lett. B **791** (2019) 130, arXiv: 1811.12446 [hep-ph] (cit. on pp. 33, 35, 42, 44).

[111] P. Baldi, P. Sadowski and D. Whiteson, *Searching for Exotic Particles in High-Energy Physics with Deep Learning*, Nature Commun. **5** (2014) 4308, arXiv: 1402.4735 [hep-ph] (cit. on pp. 35, 36, 39, 40).

[112] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, Eur. Phys. J. **C72** (2012) 1896, arXiv: 1111.6097 [hep-ph] (cit. on p. 36).

[113] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. **B641** (2006) 57, arXiv: hep-ph/0512210 [hep-ph] (cit. on p. 36).

[114] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 36).

[115] C. G. Lester and D. J. Summers, *Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders*, Phys. Lett. **B463** (1999) 99, arXiv: hep-ph/9906349 [hep-ph] (cit. on p. 36).

[116] A. Barr, C. Lester and P. Stephens, *m(T2): The Truth behind the glamour*, J. Phys. **G29** (2003) 2343, arXiv: hep-ph/0304226 [hep-ph] (cit. on p. 36).

[117] H.-C. Cheng and Z. Han, *Minimal Kinematic Constraints and m(T2)*, JHEP **12** (2008) 063, arXiv: 0810.5178 [hep-ph] (cit. on pp. 36, 37).

[118] Y. Bai et al., *Stop the Top Background of the Stop Search*, JHEP **07** (2012) 110, arXiv: 1203.4813 [hep-ph] (cit. on p. 36).

[119] D. R. Tovey, *On measuring the masses of pair-produced semi-invisibly decaying particles at hadron colliders*, JHEP **04** (2008) 034, arXiv: 0802.2879 [hep-ph] (cit. on p. 36).

[120] G. Polesello and D. R. Tovey, *Supersymmetric particle mass measurement with the boost-corrected contransverse mass*, JHEP **03** (2010) 030, arXiv: 0910.0174 [hep-ph] (cit. on p. 36).

[121] K. T. Matchev and M. Park, *A General method for determining the masses of semi-invisibly decaying particles at hadron colliders*, Phys. Rev. Lett. **107** (2011) 061801, arXiv: 0910.1584 [hep-ph] (cit. on p. 36).

[122] I. Goodfellow, Y. Bengio and A. Courville, *Deep learning*, vol. 1, 2, 2016 (cit. on p. 75).

[123] T. Chen and C. Guestrin, *XGBoost: A scalable tree boosting system*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016), URL: http://dx.doi.org/10.1145/2939672.2939785 (cit. on pp. 77, 78).

[124] J. H. Friedman, *Greedy function approximation: a gradient boosting machine*, Annals of statistics (2001) 1189 (cit. on p. 77).

[125] N. Srivastava et al., *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, Journal of Machine Learning Research **15** (2014) 1929, URL: http://jmlr.org/papers/v15/srivastava14a.html (cit. on p. 79).

[126] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, 2015, arXiv: 1502.03167 [cs.LG] (cit. on p. 79).

APPENDIX A

ML Classifiers in a Nutshell

A.1 Supervised Machine Learning

A machine learning algorithm is an algorithm that is able to learn experience from data [122]. It can be classified into supervised and unsupervised learning, depending on whether each data is associated with a target or not. Since the task in our work is supervised with two discrete target, we will focus only on binary supervised classification algorithm in this appendix. Mathematically speaking, the algorithm is trying to learn a mapping $f(\boldsymbol{x}) : \mathcal{X} \to \mathcal{Y}$, where vector $\boldsymbol{x}_i \subset \mathcal{X}$ is the i-th data sample, and $\mathcal{Y} \subseteq \{0, 1\}$ is the corresponding label.

After having learnt this mapping function (or model), we need to test its performance on new data. The performance can be evaluated by metric function, such as accuracy, which is simply the ratio of correctly predicted samples by the model. Also, it is important to emphasize that the model must be tested on unobserved data to get its generalization ability. This can be done in a simple way called "hold-out" validation. For a given data set, we randomly split it into training set and test set. As indicated by its name, training set is used to train the model, and test set to test its generalization. Usually, the model's performance differs in these two data sets. If the performance is bad for training set, we called it underfitting, which means the capacity of the model is not large enough to fit the training set. In contrast, if the model performs much worse in test set compared to training set, then it is overfitting. Overfitting is one of the major topics in machine learning, it occurs when the model's capacity is so large that it learns the local variance of training data. To avoid overfitting, some specialized techniques like regularization are applied to model to limit its capacity, but the detail is out of scope of this appendix.

In the following of this appendix, we will briefly introduce two different machine learning algorithms we used in this thesis, XGBoost and neural network. We will cover the basic ideas behind the algorithms.

A.2 XGBoost

A.2.1 Decision Tree

Before we dive into the detail of XGBoost, we first need to introduce its basic structure, the decision tree. Decision tree is a tree–like structure, it consists of a root node, multiple internal nodes, and leaf nodes. The prediction process is similar to the decision making process of human being. It starts

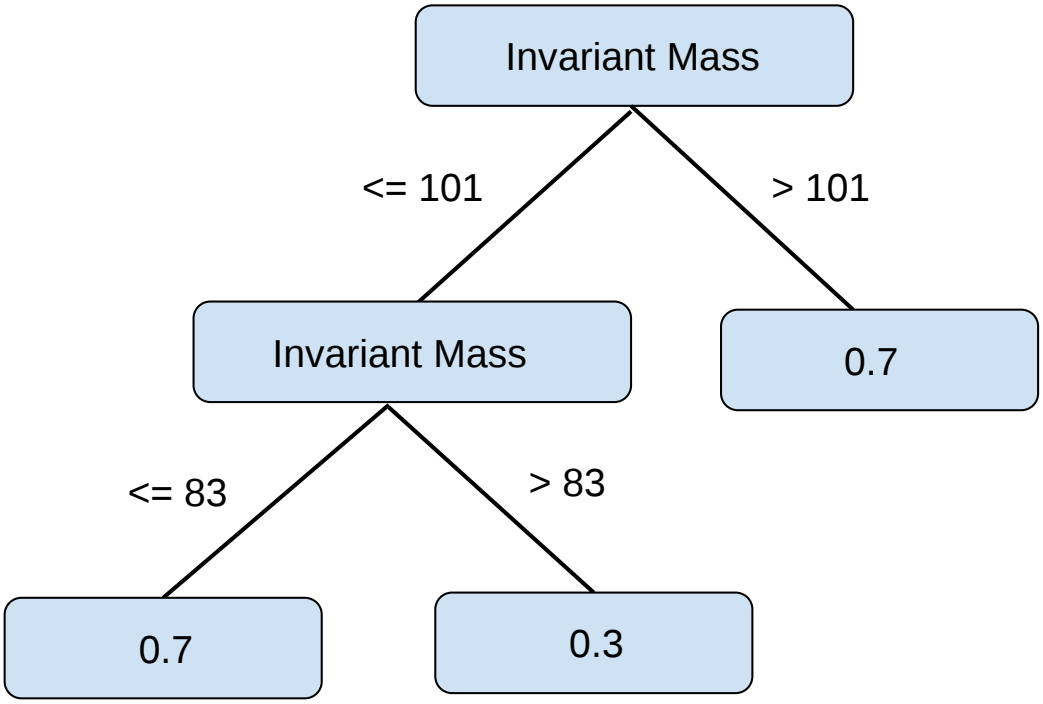

Figure A.1: An example of decision tree structure. It is built by the invariant mass of muon pairs. The number on the leaf shows the probability to be a signal.

from the root node, check its attribute and follow the conditional flow, then repeat this process until it reaches one of the leaf nodes. Then, the score or label in the leaf node corresponds to final prediction.

For example, let's consider a 2μ process $pp \to 2\mu + \not{E}_T$, where $\not{E}_T$ means invisible particles, e.g. neutrinos or DM. After training, we get a simple decision tree in Fig. A.1. If a new event, generated by the same process, has a invariant mass $m_{\mu^+\mu^-} = 95$. According to Fig. A.1, we first compare the value with the attribute in root node (the top one), if it is smaller than 101, we go left, otherwise right. By repeating this process, we finally reach a leaf node (node without any splitting) with score 0.3 on the bottom right. The score means the probability of a event to be signal. Hence, it is more likely to be a SM background according to this simple model.

It is obvious that the key part to train a decision tree is how to split a node. This includes how to choose an attribute (invariant mass) and how to split according to the chosen attribute (the conditional flow). An algorithm named ID3 (Iterative Dichotomiser 3) is more feasible to physicists, since it is based on information entropy,

$$\mathbf{Ent}(\mathcal{D}) = -\sum_{k=0}^{|\mathcal{Y}|-1} p_k \, log_2 \, p_k \, . \tag{A.1}$$

where $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}$ is the set of samples and their targets, p_k is the portion of samples labeled k. The information entropy represents the impurity of data, a smaller value means they are more likely being correctly classified. For a given attribute a and a possible splitting condition, we can split $\mathcal{D}$

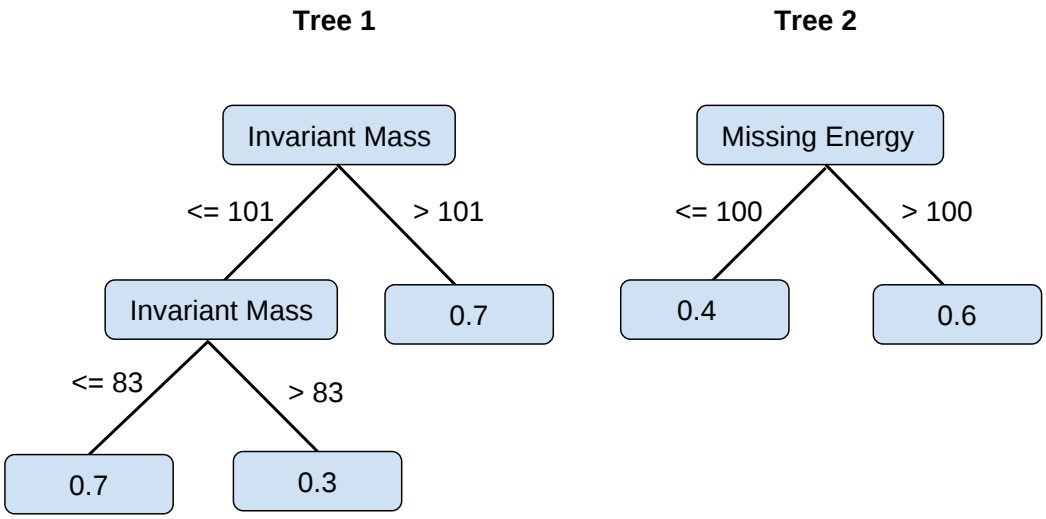

Figure A.2: An ensemble of decision trees constructed with $m_{\mu^+\mu^-}$ and $\not{E}_T$.

into two sets $\mathcal{D}_L$ and $\mathcal{D}_R$. Then we define a quantity named information gain to evaluate this splitting,

$$\textbf{Gain}(\mathcal{D}, a) = \textbf{Ent}(\mathcal{D}) - \frac{|\mathcal{D}_L|}{|\mathcal{D}|}\textbf{Ent}(\mathcal{D}_L) - \frac{|\mathcal{D}_R|}{|\mathcal{D}|}\textbf{Ent}(\mathcal{D}_R)\,. \tag{A.2}$$

Generally speaking, the information gain measures the purity improved by making this splitting. Hence, if we go through all possible attributes and splitting conditions, we can determine the current best split as the one with maximum information gain. By greedily choosing the best split, we finally obtain a decision tree.

Gradient Boosting Decision Tree

The ability of a single decision tree is usually limited, especially when the task is complicated. One way to improve its performance is by ensemble of many decision trees together, and take the sum as prediction [123].

$$\hat{y}_i = \phi(\boldsymbol{x}_i) = \sum_{v=1}^{V} f_v(\boldsymbol{x}_i)\,. \tag{A.3}$$

where $\hat{y}_i$ is the prediction and V is the number of trees. We take the same process $pp \to 2\mu + \not{E}_T$ as an example. For an event with $m_{\mu^+\mu^-} = 95$ and $\not{E}_T = 110$, we get the prediction of the decision trees in Fig. A.2 through the "vote" of 2 trees

$$\hat{y} = \frac{1}{2}(f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})) = \frac{1}{2}(0.3 + 0.6) = 0.45\,. \tag{A.4}$$

Note that we take an average of these two scores to make it probability–like.

However, naive ensemble is not good enough for some tasks, we need a more powerful algorithm named GBDT (Gradient Boosting Decision Tree) [124]. Similar to an ensemble algorithm, GBDT is also an ensemble of decision trees. But there are two major differences: firstly, GBDT generates trees

iteratively, which means that the t–th tree is dependent on previous $t-1$ trees, this is the so called boosting algorithm; secondly, GBDT generalize the process of finding the best split to minimize a predefined objective function.

In XGBoost, the objective function is written as [123],

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i,\, y_i) + \sum_v \Omega(f_v), \quad \text{where} \quad \Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 . \tag{A.5}$$

where γ and λ are hyperparameters, T is the number of leaves in the tree and w is the score in leaf node. Here l is a differentiable loss function that measures the difference between $\hat{y}_i$ and y_i. In the case of binary classification task, it can be the binary cross entropy loss,

$$l(\hat{y}_i,\, y_i) = -y_i \; log \; \hat{y}_i - (1 - y_i) \; log \; (1 - \hat{y}_i) . \tag{A.6}$$

Ω is a regularization term which penalizes the complexity of the model, such as limiting the number of leaves.

Next, we will discuss how to iteratively generate trees. Let $\hat{y}_i^{t-1}$ be the prediction of previous $t-1$ trees, then the t–th tree f_t is generated by minimize Eq. (A.5), i.e.

$$\mathcal{L}^t = \sum_i l(y_i,\, \hat{y}_i^{(t-1)} + f_t(\boldsymbol{x}_i)) + \Omega(f_t) . \tag{A.7}$$

After taking Taylor expansion up to second order and ignore constant terms, we obtain a simplified objective function at t–th iteration [123],

$$\tilde{\mathcal{L}}^t = \sum_i [g_i f_t(\boldsymbol{x}_i) + \frac{1}{2} h_i f_t^2(\boldsymbol{x}_i)] + \Omega(f_t) . \tag{A.8}$$

where $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i,\, \hat{y}_i^{(t-1)})$ and $h_i = \partial^2_{\hat{y}_i^{(t-1)}} l(y_i,\, \hat{y}_i^{(t-1)})$ are the gradients. Then after some calculations, the optimal objective value is given by [123]

$$\tilde{\mathcal{L}}^{t*} = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T . \tag{A.9}$$

where I_j is the set of samples located in leaf j. Eq. (A.9) is like the impurity score, and we can use it to find the best split by maximize the loss reduction

$$\mathcal{L}_{\text{split}} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma . \tag{A.10}$$

which is similar to maximizing the information gain in previous section. By repeating this process, we can obtain a sequence of decision trees.

Since we need to go through all possible splits, the time complexity to find a single best split is $O(n \times m)$, where n is the number of samples and m is the number of features. This is extremely time consuming, so XGBoost uses some approximate algorithms to speed it up. The details can be found in Ref. [123].

A.2.2 Neural Network

A simple example of neural network is feedforward neural network, or multilayer perceptrons (MLPs). It is called feedforward since the information flows from the input to some intermediate units, and finally to the outputs without any feedback connections. If we consider a sample with $m = 3$ features $\boldsymbol{x} = (x_1, x_2, ..., x_m)$, then a 2–layer feedforward neural network is shown in Fig. A.3. Note that, we use bold face $\boldsymbol{x}_i$ representing for i–th sample in data set, but x_i for i–th feature of a sample.

Each unit in neural network is a linear combination of units in previous layer. For example,

$$h_1 = W^h_{11}x_1 + W^h_{12}x_2 + W^h_{13}x_3 + b^h_1 \quad \Leftrightarrow \quad \boldsymbol{h} = \boldsymbol{W}^h\boldsymbol{x} + \boldsymbol{b}^h \,. \tag{A.11}$$

Here $\boldsymbol{W}^h$ and $\boldsymbol{b}^h$ are learnable parameters of a hidden layer. But linear combination is not enough for complex tasks, we need to add non–linear function to make neural network capable of learning non–linear mapping.

$$\boldsymbol{h} = \sigma(\boldsymbol{W}^h\boldsymbol{x} + \boldsymbol{b}^h) \,. \tag{A.12}$$

where σ is a non–linear function, also called activation function. A commonly used activation function in hidden layer is ReLU (Rectified Linear Unit) function,

$$\textbf{ReLU}(x) = \textbf{max}(0, x) \,. \tag{A.13}$$

As for output layer, the activation function depends on specific task. In our case, which is a binary classification problem, we use the sigmoid function. Since it maps any real number into $[0, 1]$, which is a probability–like prediction.

$$\textbf{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \,. \tag{A.14}$$

In summary, the final prediction of neural network in Fig. A.3 is,

$$\boldsymbol{h} = \textbf{ReLU}(\boldsymbol{W}^h\boldsymbol{x} + \boldsymbol{b}^h) \quad \text{and} \quad \hat{y} = \textbf{Sigmoid}(\boldsymbol{W}^o\boldsymbol{h} + \boldsymbol{b}^o) \,. \tag{A.15}$$

where $\boldsymbol{W}^o$ and $\boldsymbol{b}^o$ are the learnable parameter of the final output layer.

To determine the parameters $\boldsymbol{W}$ and $\boldsymbol{b}$ in neural network, we minimize the objective function similar to Eq. (A.5),

$$\mathcal{L} = \sum_i l(\hat{y}_i,\, y_i) + \Omega \,. \tag{A.16}$$

where Ω is also a regularization term, and $l(\hat{y}_i, y_i)$ is the same loss function in Eq. (A.6). Since the nonlinearity of neural network, the objective function becomes nonconvex. It is common to use gradient descent to find the minimum, which is taking steps in the opposite direction of gradient until reaching a local minimum. In detail, we initialize $\boldsymbol{W}$ and $\boldsymbol{b}$ to small random values, then update them according to

$$\boldsymbol{W} = \boldsymbol{W} - \eta \cdot \nabla_{\boldsymbol{W}}\mathcal{L} \quad \text{and} \quad \boldsymbol{b} = \boldsymbol{b} - \eta \cdot \nabla_{\boldsymbol{b}}\mathcal{L} \,. \tag{A.17}$$

where η is the learning rate. In practical, we use more efficient algorithms to update the parameters. They are based on gradients, but more efficient and more likely to jump out saddle point. These are called optimizers in deep learning, like `Adam` that we used in our work.

There are some other techniques that we used in our work, such as Dropout [125] to prevent overfitting and Batch Normalization [126] to speed up convergence, but they are out of scope of this

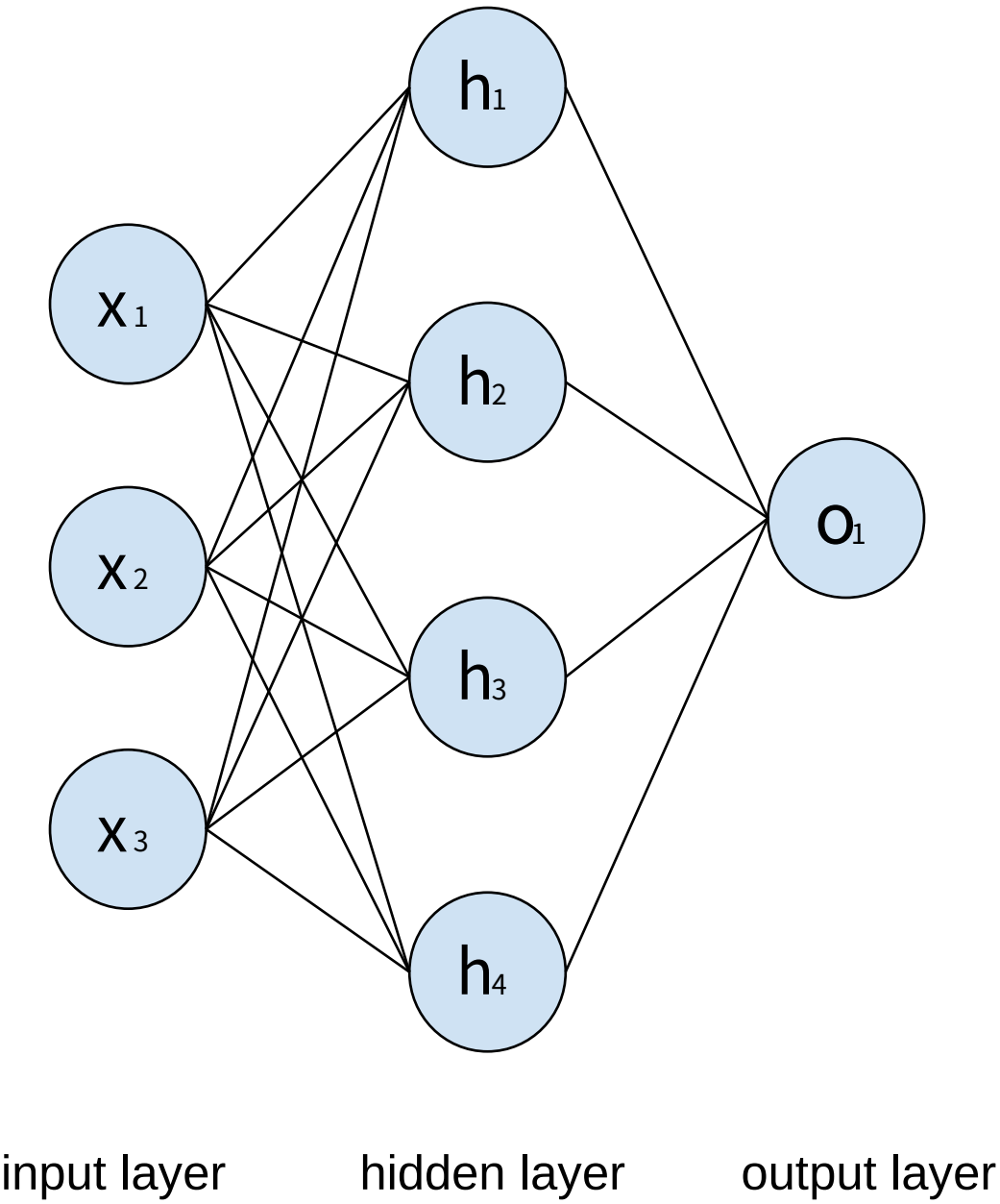

Figure A.3: The structure of a 2–layer feedforward neural network. Note that we usually do not count the input layer.

appendix.

List of Figures

List of Tables