

The application of weakly supervised learning in the search for heavy resonances at the LHC



WITS
UNIVERSITY

Nalamotse Joshua Choma

Supervised by: Prof Bruce Mellado
Co-Supervised by: Dr. Xifeng Ruan

A Thesis submitted to the Faculty of Science
School of Physics
in fulfilment
of the requirements for the degree of
Doctor of Philosophy University of the Witwatersrand

Declaration

I, Nalamotse Joshua Choma (student number 0703713G), hereby declare the contents of this thesis to be my own work unless otherwise explicitly referenced. This thesis is submitted for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. This work has not been submitted to any other university or for any other degree.



.....
Signature

03/07/2025

.....
Date

Acknowledgements

First and foremost, I would like to wholeheartedly thank my supervisor, Prof. Bruce Mellado, and my co-supervisor, Dr. Xifeng Ruan, for their unwavering support. Their persistence, drive, and extensive knowledge have helped me greatly. My heartfelt thanks go to Dr. Salah-Eddine Dahbi for his invaluable help, advice, and technical support, as well as for his patience with me in understanding the technicalities of High Energy Particle Physics. I would also like to thank the SA-CERN consortium for funding my travel expenses to CERN and the National Institute for Theoretical and Computational Sciences (NITheCS) for their continued support in funding my studies. I wish to acknowledge all my colleagues at ICCP (Institute for Collider Particle Physics). Their kind help and support made this work possible, especially Gaogalalwe Mokgatitswane.

Most importantly, constant thanks are due to my family: my mother Rebecca Choma, my brother Tumelo Choma, my uncle Makgatle Choma, my partner Pontsho Kekae, and the rest of my family and friends, especially Tebogo Molefe, Ben Mahudu, and Daniel Choshi, not forgetting the members of the ZCC student fellowship. They all played a crucial role in getting me this far. You have encouraged me, supported me spiritually, morally and emotionally, and kept me in your prayers for my ultimate success. This thesis is a testimony of your unconditional love and encouragement. Last but not least, to the God of Mount Zion, **“Dithetho dia a go swanela!”**.

Dedication

This thesis is dedicated to my late grandmother, Mankhedi Betty Choma, kgadi ya babina tlou. May your soul continue to rest in peace tlou ya mamphela. I will be forever grateful for the values you instilled in me. Your teachings will always live on in me. The sacrifices you made to ensure that I am a better person will never go unnoticed. In your favorite words, *'tsena sekolo ramogolo gore o se ke wa tlo ja ditšhila tša dithaka tša gago.'* Even if you never witnessed any of my graduations after high school, this is for you Pebetsi'a Hlabirwa le Mahlako.



Publications

Some ideas and figures have appeared previously in the following publications:

Journals

1. S.-E Dahbi, **J. Choma**, B. Mellado, G. Mokgatitswane, X. Ruan, B. Lieberman, and T. Celik, "Machine learning approach for the search of resonances with topological features at the Large Hadron Collider" [1].

Conferences

1. **J. Choma**, S.-E Dahbi, B. Mellado, and X. Ruan, "The use of Semi-supervision in the search for heavy resonances with the $Z\gamma$ final state at the LHC". Published in the South African Institute of Physics (SAIP2021) proceedings, ISBN: 978-0-620-97693-0 [2].
2. **J. Choma**, S.-E Dahbi, G. Mokgatitswane, X. Ruan, and B. Mellado, "The search for resonances with topological requirements with the $Z\gamma$ final state at the LHC", Pan-African Astro-Particle and Collider Physics Workshop, 2022, Submitted for review and publication.
3. S.-E Dahbi, **J. Choma**, B. Mellado, G. Mokgatitswane, X. Ruan, and B. Lieberman, "Machine learning approach for the search of resonances with topological features at the Large Hadron Collider". Published in the South African Institute of Physics (SAIP2021) proceedings, ISBN: 978-0-620-97693-0 [3].
4. B. Lieberman, **J. Choma**, S.-E Dahbi, B. Mellado, and X. Ruan, "An investigation of over-training within semi-supervised machine learning models in the search for heavy resonances at the LHC". Published in the South African Institute of Physics (SAIP2021) proceedings, ISBN: 978-0-620-97693-0 [4].
5. G. Mokgatitswane, **J. Choma**, S.-E Dahbi, X. Ruan, and B. Mellado, "Study of systematic uncertainties and spurious signals of resonant $H \rightarrow Z\gamma$ production at ATLAS Experiment". Published in the South African Institute of Physics (SAIP2022) proceedings, ISBN: 978-0-6397-4426-1 [5].

Abstract

The discovery of the Higgs boson at the Large Hadron Collider by the ATLAS and CMS experiments has made the search for new physics beyond the Standard Model a priority in the field of High Energy Particle Physics. New resonances have yet to be discovered using inclusive and model-dependent searches, which means they may be driven by subtle topologies. Rapid improvements in Machine Learning techniques have led to their increasing application in High Energy Particle physics. Unlike supervised learning, which is known to assume full knowledge of the underlying model, semi-supervised learning, in particular weakly supervised learning, allows the extraction of new information from data with partial knowledge. The goal of this study is to set up searches for heavy resonances at the electroweak scale with topological requirements performed in both inclusive and exclusive regions of phase-space tailored to a particular production mode. These resonances could be generated with different production mechanisms. In this work, we describe search procedures based on weakly supervised learning applied to mixed samples and used to search for resonances with little or no prior knowledge of the production mechanism. This approach has the advantage that sidebands or control regions can be used to effectively model backgrounds without relying on models. The effectiveness of this method is measured by the production of the Standard Model Higgs boson, which decays into a pair of photons in both inclusive and exclusive regions of phase-space at the LHC. Having confirmed the ability of the method to extract various Standard Model Higgs boson signal processes, the search for new phenomena in high mass final states will be set up at the LHC. Subsequently, the approach is used in the search for new resonances in the $Z\gamma$ final state with $Z \rightarrow e^+e^-$ or $Z \rightarrow \mu^+\mu^-$, using the Monte Carlo simulated signal samples for 139 fb^{-1} of integrated luminosity for Run 2 collected at the LHC. The weakly supervised learning approach is implemented and compared to the performance of the fully supervised approach, which is then used to calculate the production limit for Higgs-like particles for $Z\gamma$ where the significance of the signal is maximal.

Contents

Declaration	i
Acknowledgements	ii
Dedication	iii
Publications	iv
Abstract	iv
Contents	vi
List of Figures	xiii
List of Tables	xv
List of Abbreviations	xvii
List of Symbols	1
1 Introduction	2
2 The Standard Model and Beyond	4
2.1 The Standard Model of Particle Physics	4
2.1.1 Electroweak Theory	5
2.1.1.1 Quantum Electrodynamics	6
2.1.1.2 Quantum Chromodynamics	6
2.1.2 Gauge Bosons and Interactions	7
2.1.3 Elementary Fermions	8
2.2 The BEH Mechanism and The Higgs Boson	8
2.3 Beyond the Standard Model	9
2.3.1 The Madala Hypothesis	10
2.3.1.1 Testing the hypothesis in the multi-lepton final states	13
3 The Large Hadron Collider and the ATLAS Detector	16
3.1 The Large Hadron Collider	16
3.2 The ATLAS Detector	19

3.2.1	ATLAS Coordinate System	20
3.2.2	Inner Detector	22
3.2.2.1	Pixel Detector	22
3.2.2.2	Semiconductor Tracker	22
3.2.2.3	Transition Radiation Tracker	22
3.2.3	Calorimeters	23
3.2.3.1	Electromagnetic Calorimeter	24
3.2.3.2	Hadronic Calorimeters	24
3.2.4	Muon spectrometers	25
3.3	ATLAS Trigger System	26
3.4	Event Reconstruction in ATLAS	27
3.4.1	Track Reconstruction	28
3.4.2	Electron Reconstruction	29
3.4.3	Jet Reconstruction	29
3.4.4	Muon Reconstruction	29
3.4.5	Photon Reconstruction	30
3.4.6	Missing Transverse Energy	30
4	Semi-supervised Learning-based search of heavy resonances decay to $\gamma\gamma$ with topological features	32
4.1	Introduction	32
4.2	Application of Machine Learning in HEP	33
4.2.1	Random Forest	34
4.2.2	Deep Learning	36
4.2.3	Fully Supervised Learning	37
4.2.4	Semi-supervised Learning	38
4.2.4.1	Weakly Supervised Learning	38
4.2.4.2	Supervised learning with unlabelled data	39
4.3	Dataset and event selection	40
4.3.1	Kinematic features	42
4.4	The DNN configuration	45
4.4.1	Backpropagation	47
4.5	Fully Supervised Learning Classification	47
4.6	Semi-supervised learning classification	48
4.6.1	Supervised classification with unlabelled signals	48
4.6.2	Effect of signal event yield on the event classification	50
4.6.3	Weakly supervised learning classification	50
4.6.4	Weakly supervised learning with topological requirements	51
4.7	Discussion and conclusions	56
5	Application of Semi-supervision in the search for resonances with the $Z\gamma$ final state at the LHC	58
5.1	Introduction	58
5.2	Data and Simulated Samples	59
5.2.1	Data	59
5.2.2	Backgrounds	59

5.3	Object Reconstruction and Overlap Removal	59
5.4	Data and Kinematic Features Selection	60
5.4.1	Data Processing	64
5.4.2	Random Forest Configuration	67
5.4.3	The Deep Neural Networks Configuration	67
5.5	Performance	70
5.5.1	Significance	74
5.6	Spurious Studies	75
5.7	Signal Parameterization	78
5.8	Discussion and Conclusions	81
6	Conclusions	83
	Appendices	84
A	Correlation Matrix	86
A.1	Inclusive	87
A.2	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	89
A.3	$N_{jets} \geq$ and $60 < m_{jj} < 120 \text{ GeV}$	91
B	Random Forest Response Distributions	93
B.1	Inclusive	94
B.1.1	200 GeV	94
B.1.2	300 GeV	94
B.1.3	500 GeV	95
B.2	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$	96
B.2.1	150 GeV	96
B.2.2	200 GeV	96
B.3	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$	97
B.3.1	150 GeV	97
B.3.2	200 GeV	97
B.4	$N_{jets} \geq 2$ and $60 < m_{jj} < 120 \text{ GeV}, Wh$	98
B.4.1	150 GeV	98
B.4.2	200 GeV	98
B.5	$N_{jets} \geq 2$ and $60 < m_{jj} < 120 \text{ GeV}, Zh$	99
B.5.1	150 GeV	99
B.5.2	200 GeV	99
C	DNN Response Distributions	100
C.1	Inclusive	101
C.1.1	200 GeV	101
C.1.2	300 GeV	101
C.1.3	500 GeV	102
C.2	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$	102
C.2.1	150 GeV	102
C.2.2	200 GeV	103
C.3	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$	103

C.3.1	150 GeV	103
C.3.2	200 GeV	104
C.4	$N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh	104
C.4.1	150 GeV	104
C.4.2	200 GeV	105
C.5	$N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh	105
C.5.1	150 GeV	105
C.5.2	200 GeV	106
C.6	VBF	106
C.6.1	150 GeV	106
C.6.2	200 GeV	107
C.6.3	300 GeV	107
C.6.4	500 GeV	108
D	Significance graphs	109
D.1	Inclusive	110
D.1.1	200 GeV	110
D.1.2	300 GeV	110
D.1.3	500 GeV	111
D.2	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$	112
D.2.1	150 GeV	112
D.2.2	200 GeV	112
D.3	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$	113
D.3.1	150 GeV	113
D.3.2	200 GeV	113
D.4	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$	114
D.4.1	150 GeV	114
D.4.2	200 GeV	114
D.5	$N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh	115
D.5.1	150 GeV	115
D.5.2	200 GeV	115
D.6	$N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh	116
D.6.1	150 GeV	116
D.6.2	200 GeV	116
D.7	VBF	117
D.7.1	150 GeV	117
D.7.2	200 GeV	117
D.7.3	300 GeV	118
D.7.4	500 GeV	118
E	ROC Curves	119
E.1	Inclusive	120
E.2	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	121
E.3	$N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV category	122
E.4	VBF Category	123

F	Spurious Signal	124
F.1	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	125
F.2	$N_{jets} \geq 2$ and $75 \leq m_{jj} \leq 105 \text{ GeV}$	128
F.3	$N_{jets} \geq 1$	131
G	Signal Parameterization	134
	Bibliography	134

List of Figures

2.1	Schematic illustration of the Standard Model particles showing the three generations of matter/fermions, gauge bosons and the Higgs boson [26].	5
2.2	The primary production mechanisms for the SM Higgs boson at high energy pp collisions.	10
2.3	Dilepton invariant mass distribution of ATLAS data (dots) and SM background (red line) compared with the BSM signal (blue line) for events requiring a pair of oppositely charged leptons of a different flavor ($e^\pm\mu^\pm$) with at least one b -tagged jet. [56].	13
2.4	The (negative 2 times logarithm of the) profile likelihood ratios for each of the individual fit outcomes, overlaid on that of their combination [6].	14
3.1	An illustration of the CERN accelerator complex [61].	17
3.2	The recorded ATLAS integrated luminosity in different years [66].	18
3.3	The schematic illustration of the ATLAS detector [62].	19
3.4	The ATLAS coordinate system showing the collision point in the Cartesian plane xyz with polar and azimuthal angles [62].	20
3.5	(a) The cutway view of the ATLAS Inner Detector system [68], (b) The Inner Detector of ATLAS [69].	21
3.6	The cutway view of the ATLAS calorimeter system [62].	23
3.7	The ATLAS Muon Spectrometer [62].	25
3.8	Schematic representation of the ATLAS trigger and data acquisition systems after Run 2 [78].	27
3.9	Schematic representation of different particles interaction with the ATLAS subsystem [80].	28
4.1	A schematic illustration of a Random Forest comprised of N number of decision trees.	34
4.2	Schematic representation of supervised learning.	37
4.3	The spectrum of the invariant di-photon mass at 13 TeV of the proton-proton collision corresponds to an integrated luminosity of 36.1 fb^{-1} . The cross sections times the branching ratio of the di-photon decays are used to standardize the different production modes of the Higgs boson samples of the SM	41

4.4	Signal and background distributions of the object multiplicity: (a) number of reconstructed jets per event, (b) number of b -jets per event identified by the b -tagging algorithm, and (c) number of reconstructed leptons per event	42
4.5	Signal and background distributions of (a) pseudorapidity of the leading jet, (b) pseudorapidity of the sub-leading jet, (c) transverse momentum of the leading jet, (d) transverse momentum of the sub-leading jet, (e) pseudorapidity separation of the leading and sub-leading jet, and (f) invariant mass of the leading and sub-leading system	44
4.6	Signal and background distributions of: (a) reconstructed missing transverse energy, (b) Azimuthal angle separation between the leading jets and the missing transverse energy, (c) Azimuthal angle separation between the di-photon system and the missing transverse energy (d) Azimuthal angle separation between the di-photon system and the leading jets.	45
4.7	The DNN architecture configuration for the classification of events as background or signal in this study.	46
4.8	Fully supervised learning DNN response distributions for the Higgs production, by: (a) gluon-gluon fusion and (b) VBF.	48
4.9	Fully supervised learning DNN response distributions for the SM Higgs boson associated production, with: (a) W vector boson, (b) Z vector boson or (c) a top quark pair.	49
4.10	ROC curves of DNN classification with fully supervised learning for different Higgs production mechanisms.	50
4.11	Results of fully supervised learning with unlabelled signals. Left: DNN response distributions. Right: ROC curves, for different Higgs production modes.	51
4.12	ROC curves from fully supervised learning with unlabelled signals, (a) VBF, (b) Wh , (c) Zh and (d) $t\bar{t}h$ are normalized to different proportions of the expected ggF production mechanism events.	52
4.13	ROC curves from weakly supervised learning using the DNN classifier for different Higgs production mechanisms. Signals and background are normalized to their cross-sections.	53
4.14	Weakly supervised learning results with VBF topological selections. For the VBF production mode, the diagram on the left shows a two-dimensional scan of the signal significance with respect to m_{jj} and $\Delta\eta_{jj}$. The expected signal significance is represented by the z axis. The ROC curves for the VBF production mechanism with various learning methods are shown in the graph on the right (see text). Weakly supervised learning results are not as good as unlabelled or fully supervised at background rejection. . .	54
4.15	The solid lines show the ROC curves in inclusive Wh final states using different learning approaches. The leptonic and hadronic Wh topology requirements with weakly supervised learning are shown with dashed lines.	55
4.16	The solid lines show the ROC curves for a range of different learning methods in inclusive $t\bar{t}h$ final states. The dashed lines show the weakly supervised learning with $t\bar{t}h$ topological requirements.	56

5.1	Invariant mass with an injection of signal for the following categories: (a) Inclusive, (b) $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$, (c) $N_{jets} \geq 2$; $60 < m_{jj} < 120 \text{ GeV}$ and (d) $N_{jets} \geq 2$; $\Delta\eta_{jj} > 2$; $m_{jj} > 300 \text{ GeV}$	61
5.2	Correlation matrix of the implemented features/kinematics for the 200 GeV background VBF category.	62
5.3	Correlation matrix of the implemented features/kinematics for the VBF signal for 200 GeV.	63
5.3	The signal and background distribution of: (a) pseudorapidity of the leading jet, (b) pseudorapidity of the sub-leading jet, (c) pseudorapidity of the dijet system, (d) invariant dimass jet of the leading and sub-leading system, (e) Azimuthal angular separation between the forwards jets and missing transverse energy and (f) Azimuthal angular separation between the leading jets and missing transverse energy.	64
5.3	The signal and background distribution of: (g) Azimuthal angular separation between the soft jets and missing transverse energy, (h) Azimuthal angular separation between the $Z\gamma$ system and missing transverse energy, (i) missing transverse energy, (j) Azimuthal angle of the missing transverse energy, (e) missing significant energy and (f) the number of jets.	65
5.3	The signal and background distribution of: (m) number of central jets, (n) relative pt of a photon, (o) Azimuthal angular separation between the $Z\gamma$ system and the dijet system and (p) Azimuthal angular separation between the $Z\gamma$ system and the leading jet.	66
5.4	schematic representation of cross-validation.	67
5.5	Model loss as a function of epochs for: (a) DNN with an implementation of the early stopping parameter and (b) DNN without the early stopping parameter	69
5.6	The DNN response distributions: (a) Weakly supervised learning for $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category, (b) Fully supervised learning for $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category, (c) Weakly supervised learning for $N_{jets} \geq 2$; $60 < m_{jj} < 120 \text{ GeV}$ category and (d) Fully supervised learning for $N_{jets} \geq 2$; $\Delta\eta_{jj} > 2$; $m_{jj} > 300 \text{ GeV}$ category.	71
5.7	The RF response distributions: (a) Weakly supervised learning for $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category, (b) Fully supervised learning for $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category, (c) Weakly supervised learning for $N_{jets} \geq 2$; $60 < m_{jj} < 120 \text{ GeV}$ category and (d) Fully supervised learning for $N_{jets} \geq 2$; $\Delta\eta_{jj} > 2$; $m_{jj} > 300 \text{ GeV}$ category.	72
5.8	The ROC curves for: (a) $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$, (b) $N_{jets} \geq 2$; $\Delta\eta_{jj} > 2$; $m_{jj} > 300 \text{ GeV}$	72
5.9	Fitting on the background MC template VBF 150 GeV.	77
5.10	Fitting on the background template for VBF 150 GeV, with the S+B function.	77
5.11	The $N_{spurious}/\delta N_{spurious}$ as a function of mass for each functional form for VBF 150 GeV.	78
5.12	Results of individual fit of DSCB parameters as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for VBF inclusive: (a) 200 GeV (b) 300 GeV and (c) 500 GeV.	80

5.13	Parameterization of: (a) μ_{CB} and (b) σ_{CB} as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for VBF	81
A.1	Correlation matrix of the implemented features/kinematics for the 200 GeV background in the inclusive category.	87
A.2	Correlation matrix of the implemented features/kinematics for the ggF signal for 200 GeV in the inclusive category.	88
A.3	Correlation matrix of the implemented features/kinematics for the 200 GeV background in the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category.	89
A.4	Correlation matrix of the implemented features/kinematics for the Wh signal for 200 GeV in the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category.	90
A.5	Correlation matrix of the implemented features/kinematics for the 200 GeV background in the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV category. . .	91
A.6	Correlation matrix of the implemented features/kinematics for the Zh signal for 200 GeV in the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV category. .	92
B.1	RF Response Distribution for the inclusive category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	94
B.2	RF Response Distribution for the inclusive category for 300 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	94
B.3	RF Response Distribution for the inclusive category for 500 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	95
B.4	RF Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	96
B.5	RF Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$ category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	96
B.6	RF Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	97
B.7	RF Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	97
B.8	RF Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	98
B.9	RF Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	98
B.10	RF Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	99
B.11	RF Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	99

C.1	DNN Response Distribution for the inclusive category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	101
C.2	DNN Response Distribution for the inclusive category for 300 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	101
C.3	DNN Response Distribution for the inclusive category for 500 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	102
C.4	DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	102
C.5	DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$ category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	103
C.6	DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	103
C.7	DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	104
C.8	DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	104
C.9	DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	105
C.10	DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	105
C.11	DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	106
C.12	DNN Response Distribution for the VBF category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	106
C.13	DNN Response Distribution for the VBF category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	107
C.14	DNN Response Distribution for the VBF category for 300 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	107
C.15	DNN Response Distribution for the VBF category for 500 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	108
D.1	DNN Response Distribution for the inclusive category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	110
D.2	DNN Response Distribution for the inclusive category for 300 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	110
D.3	DNN Response Distribution for the inclusive category for 500 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	111

D.4	DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$, Wh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	112
D.5	DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$, Wh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	112
D.6	DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$, Zh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	113
D.7	DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$, Zh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	113
D.8	DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$, Zh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	114
D.9	DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$, Zh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	114
D.10	DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	115
D.11	DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	115
D.12	DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	116
D.13	DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	116
D.14	DNN Response Distribution for the VBF category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	117
D.15	DNN Response Distribution for the VBF category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	117
D.16	Significance for the VBF category for 300 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	118
D.17	DNN Response Distribution for the VBF category for 500 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.	118
E.1	ROC curves for the inclusive category for the following mass points: (a) 200 GeV, (b) 300 GeV and (c) 500 GeV learning.	120
E.2	ROC curves for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category for the following mass points: (a) Wh 150 GeV, (b) Wh 200 GeV, (c) Zh 150 GeV and (d) Zh 200 GeV.	121

E.3	ROC curves for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV category for the following mass points: (a) Wh 150 GeV, (b) Wh 200 GeV, (c) Zh 150 GeV and (d) Zh 200 GeV.	122
E.4	ROC curves for the VBF category for the following mass points: (a) 150 GeV, (b) 200 GeV, (c) 300 GeV and (d) 500 GeV learning	123
F.1	Fitting on the background template for VH Leptonic 150 GeV, with S+B function.	126
F.2	The $N_{spurious}/\delta N_{spurious}$ as a function of mass for each functional form for VH Leptonic 150 GeV.	127
F.3	Fitting on the background template for VH Hadronic 150 GeV, with S+B function.	129
F.4	The $N_{spurious}/\delta N_{spurious}$ as a function of mass for each functional form for VH Hadronic 150 GeV.	130
F.5	Fitting on the background template for ttH 150 GeV, with S+B function.	132
F.6	The $N_{spurious}/\delta N_{spurious}$ as a function of mass for each functional form for ttH 150 GeV.	133
G.1	Results of individual fit of DSCB parameters as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for Wh (metsig): (a) 150 GeV and (b) 200 GeV.	136
G.2	Results of individual fit of DSCB parameters as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for Zh (metsig): (a) 150 GeV and (b) 200 GeV.	136
G.3	Results of individual fit of DSCB parameters as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for Wh (mjj): (a) 150 GeV and (b) 200 GeV.	137
G.4	Results of individual fit of DSCB parameters as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for Zh (mjj): (a) 150 GeV and (b) 200 GeV.	138

List of Tables

2.1	ATLAS and CMS results utilized to constrain the relevant parameters of the Madala hypothesis [7].	11
2.2	A summarized results of the SM+BSM fit and their combination [6]. . .	14
4.1	The expected events are normalized to the number of entries for each process.	41
4.2	Relative contribution of each signal to the performance of the DNN classifier, for fully supervised learning with unlabelled signal (Section 4.6.1). Signals are normalized to different proportions of the expected events from the ggF production mechanism.	51
4.3	Comparison of the integral of ROC curve AUCs from different learning methods.	54
4.4	VBF, Wh and $t\bar{t}h$ ROC integrals with different topological requirements, compared to inclusive weakly supervised learning.	56
5.1	Events yield for the considered processes and signal injection rates for VBF, ggF, Zh and Wh signal samples at different mass points.	60
5.2	The parameters used to configure the Random Forest in Python-based package scikit-learn [217].	67
5.3	The hyper-parameters used to configure different DNN models.	68
5.4	A comparison of the performance of DNNs and RFs in the form of ROC AUC based on weakly supervised learning.	73
5.5	A comparison of the performance of DNNs and RFs in the form of ROC AUC based on fully supervised learning.	73
5.7	Maximum significance for various background rejections from the DNN response distribution with weakly supervised learning.	75
5.6	Maximum significance for various background rejections from the DNN response distribution with fully supervised learning.	75
5.8	Fitted spurious signal yield at 150 GeV for VBF phase-space. The spurious signal yield is normalized to 140 fb^{-1} . Category-0 represents no background rejection, whereas Category-1 to Category-6 represent 10%, 20%, 30%, 40% and 50% background rejection.	76
5.9	Summary of μ_{CB} and σ_{CB} values for VBF signal samples at $N_{jets} \geq 2$, $m_{jj} > 300 \text{ GeV}$ and $\Delta\eta_{jj} > 2$, after background rejection.	79
F.1	Fitted spurious signal yield at 150 GeV for VH Leptonic phase space, $SE_T^{miss} > 2.5 \sqrt{\text{GeV}}$	125

F.2	Fitted spurious signal yield at 150 GeV for VH Hadronic phase space. . .	128
F.3	Fitted spurious signal yield at 150 GeV for VH Leptonic phase space. . .	131
G.1	Summary of μ_{CB} and σ_{CB} values for Wh 150 GeV signal sample at $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ after background rejection.	134
G.2	Summary of μ_{CB} and σ_{CB} values for Zh 150 GeV signal sample at $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ after background rejection.	135
G.3	Summary of μ_{CB} and σ_{CB} values for Wh 150 GeV signal sample at $N_{jets} \geq 2$ and $60 < m_{jj} < 120 \text{ GeV}$ after background rejection.	135
G.4	Summary of μ_{CB} and σ_{CB} values for Zh 150 GeV signal sample at $N_{jets} \geq 2$ and $60 < m_{jj} < 120 \text{ GeV}$ after background rejection.	137

List of Abbreviations

AI	Artificial Intelligence
ALICE	A Large Ion Collider Experiment
ANN	Artificial neural network
ATLAS	A Toroidal LHC Apparatus
AUC	Area Under the Curve
BDT	Boosted Decision Tree
BEH	Brout-Englert-Higgs
BSM	Beyond the Standard Model
CERN	European Organization for Nuclear Research
CMS	Compact Muon Solenoid
CNN	Convolutional Neural Networks
CSC	Cathode Strip Chambers
DL	Deep Learning
DT	Decision Tree
ECal	Electromagnetic Calorimeter
EW	Electroweak
FCal	Forward Calorimeter
GeV	Giga Electronvolt
ggF	Higgs production via gluon gluon fusion
HCal	Hadronic Calorimeter
HEC	Hadronic end-cap Calorimeter
HEPP	High Energy Particle Physics
HRI	Harish-Chandra Research Institute
IBL	Insertable B-layer
ICPP	Institute for Collider Particle Physics
ID	Inner Detector

KNN	K-Nearest Neighbour
LAr	Liquid Argon
LHC	Large Hadron Collider
LO	Leading Order
MC	Monte Carlo
MDT	Monitored Drift Tubes
ML	Machine Learning
MLP	Multilayer Perceptron
MS	Muon Spectrometer
NLO	Next-to-leading Order
RF	Random Forest
RPC	Resistive Plate Chambers
PS	Proton Synchrotron
PSB	Proton Synchrotron Booster
ROC	Receiver Operator Characteristic
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
SCT	Semi-conductor Tracker
SM	Standard Model
SULU	Supervised Learning using Labelled and Unlabelled
TGC	Thin Gap Chambers
TMVA	Toolkit for Multi-Variate Data Analysis
SPS	Super Proton Synchrotron
SUSY	Supersymmetry
SVM	Support Vector Machine
TRT	Transition Radiation Tracker
$t\bar{t}h$	Higgs production via top quark fusion
VBF	Higgs production via vector boson fusion
Vh	vector boson
Wh	Higgs production via W boson associated production
Wits	University of the Witwatersrand
Zh	Higgs production via Z boson associated production

List of Symbols

fb^{-1}	Inverse femtobarns
$W^{\pm}$	W bosons
Z	Z boson
ψ	Charged fermion
η	Pseudorapidity
b	Bottom quark
c	Charm quark
d	Dark quark
h	Standard Model Higgs boson
H	Heavy scalar particle
S	Higgs-like scalar
χ	Dark matter candidate
v	The vacuum expectation value
$\sqrt{s}$	Center-of-mass energy
σ	Cross-section of protons collisions
$\mathcal{L}$	Luminosity
α	Learning rate

Chapter 1

Introduction

Scientists benefit from advances in technology and scientific knowledge in their efforts to understand and answer questions like: *What is the universe made of? How does it interact?* Among these advances is the Large Hadron Collider (LHC), which scientists use to collide particles. Particle physicists can use these collisions at high energies to reproduce the conditions that existed seconds after the Big Bang. The remnants of these collisions are being studied to answer the previous two fundamental questions. Particle physicists have worked hard to investigate and discover the underlying physical laws that govern the universe. To understand the interaction of subatomic particles, they are guided by a strong theory known as the Standard Model (SM). Although this theory has been extremely successful in making predictions, there are still certain inconsistencies that the theory cannot explain. Therefore, several possible extensions of the SM have been presented to address the shortcomings of this theory. These extensions are referred to as Beyond the Standard Model (BSM) theories.

The large amount of data generated by the LHC requires significant resources for processing and storage. Currently, the LHC generates petabytes of data per year. The vast majority of the data generated by the collisions produces particles of low interest in the form of background, while a small percentage of the data produces particles of interest in the form of a signal. Therefore, the signal events are difficult to identify since they are too tiny to be observed in real-time and can quickly decay into other particles. Complex statistical approaches are required to distinguish the signal from the background. Among other approaches, machine learning (ML) techniques play an important role in processing the LHC data. The discovery of the Higgs boson (see Chapter 2.2) was an important contribution from them. In particular, Deep Learning (DL) algorithms can learn from highly complex data and identify nonlinear feature combinations. Therefore, they are a suitable tool to study large datasets of high energy particle physics (HEPP).

The search for new resonances is very important towards the goal of unveiling new physics BSM. Thus, the absence of signals in inclusive searches implies that if the new resonances are hidden in the data, the fraction of the signal is very small. Machine learning, however, can help mitigate this problem. Machine learning techniques such as fully supervised learning have proven successful in exploring corners of the phase-space

of interest. The phase-spaces where the SM boson was observed would be an ideal example. However, this approach needs a well-defined model and this may not happen in the real data. For this reason, the use of techniques such as semi-supervised learning, in particular weakly supervised learning, is more attractive for this type of search since it can bring more freedom to the requirements of the signal models. This is based on the advantage that the sideband is able to detect topological differences between the potential signal and the sidebands. In a search for resonance, the sideband is treated as the absence of the signal, thus it can be treated as pure background. So it can be used for training versus the signal region, where it is a mixture of signal and background. However, given the complexity of the phase-space under study and the way it is populated by background and potential signals, this approach could be considered ineffective. Therefore, to mitigate this problem, this study focuses on several predefined regions to narrow down the phase-space of the search.

Several anomalies were detected in both the ATLAS and CMS experiments. These anomalies are based on the discrepancies in the production of leptons [6–16] and they are further accommodated by the formulated hypothesis discussed in detail in Chapter 2.3.1. This serves as a motivation for the search for Higgs-like resonances with mass around the electroweak (EW) scale, where SM backgrounds are large. In this case, topological requirements play a crucial role in hiding the backgrounds. The approach is further used to search for heavy resonances decaying to $Z\gamma$ final states, which describes a search for heavy resonances where $Z \rightarrow e^+e^-$ or $Z \rightarrow \mu^+\mu^-$. The data used in this analysis correspond to the integrated luminosity of 139 fb^{-1} . Moreover, the study is an ongoing analysis whose sole purpose is to initiate the search for heavy resonances in the EW scale. Therefore, it is performed with topological requirements using ML techniques.

This thesis is structured as follows: The introduction to the Standard Model and the possible physics beyond it are covered in Chapter 2. The LHC and the ATLAS detector are briefly described in Chapter 3, while Chapters 4 and 5 focus on machine learning-based analysis of $\gamma\gamma$ and $Z\gamma$ final states, respectively. Finally, Chapter 6 summarizes the work presented in this thesis.

Chapter 2

The Standard Model and Beyond

Particle physics is a discipline of physics that deals with elementary particles and their properties. These elementary particles have no substructure and are considered fundamental components of all objects in the universe [17]. In addition, they describe three of the four interactions observed in nature. Their discoveries have been made over time, with the 2012 discovery of the Higgs boson by the ATLAS [18] and CMS [19] collaborations, forming the final piece of the Standard Model. This Chapter outlines the theoretical structure of the SM.

2.1 The Standard Model of Particle Physics

The SM is considered the most successful theory in particle physics. This theory was developed in the 1970s and is responsible for describing all fundamental particles and their interactions at the subatomic level [20–23]. It has been used to predict the existence and properties of various particles that have been experimentally confirmed [24]. The theory is composed of two components, namely the electroweak force and quantum chromodynamics (QCD). The electroweak theory describes the interactions via the electromagnetic and the weak force, while the latter is the theory of the strong nuclear force. These theories are gauge field theories. They describe the interaction between particles and are governed by the gauge principle, which allows to obtain a Lagrangian from a symmetry.

The SM consists of 17 fundamental particles, of 12 fermions and five bosons, which are divided into categories [25]. Their categories are based on intrinsic angular momentum or spin. These particles are organized as shown in Figure 2.1. The fermions are the fundamental building blocks of matter, while the bosons serve as mediators of interactions. These fermions are divided into three generations, with the first generation consisting of lighter particles. The particles of the last two generations have identical quantum numbers, and higher masses compared to the particles of the first generation. These high masses make them unstable, so they can decay into generation 1 particles.

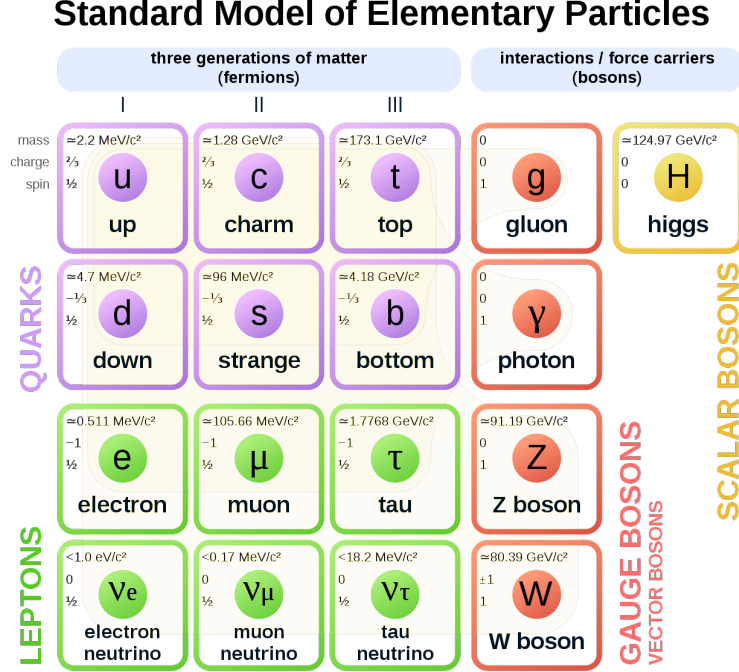


Figure 2.1: Schematic illustration of the Standard Model particles showing the three generations of matter/fermions, gauge bosons and the Higgs boson [26].

2.1.1 Electroweak Theory

The unification of quantum electrodynamics (QED) and weak theory is known as electroweak theory. The gauge symmetry $SU(2)_L \times U(1)_Y$ describes this unification, where $SU(2)_L$ is the symmetry of the weak interactions and $U(1)$ is the symmetry of QED [27]. The weak and electromagnetic interactions between leptons and quarks are described by this theory [21–23]. This symmetry group is very complicated, in contrast to the QCD theory, which is based on the symmetry group $SU(3)_C$, the electroweak theory is based on the symmetry group $SU(2)_L \times U(1)_Y$, which describes the weak force [27]. Moreover, QCD theory explains the strong interaction between colored quarks.

The gauge fields of the electroweak theory which are physically observable are the photon field A , the neutral Z boson field Z^0 and the two charged W boson fields $W^\pm$. These gauge fields are given by:

$$A = \sin\theta_W W^3 + \cos\theta_W B, \quad (2.1)$$

$$Z^0 = \cos\theta_W W^3 - \sin\theta_W B, \quad (2.2)$$

$$W^\pm = \sqrt{1/2}(W^1 \mp iW^2), \quad (2.3)$$

where θ_W represents the Weinberg angle and is defined in terms of weak isospin and weak hypercharge as:

$$\tan\theta_W = \frac{g_1}{g_2}. \quad (2.4)$$

In order to understand this we need to explain QED and QCD.

2.1.1.1 Quantum Electrodynamics

The electromagnetic interaction between charged particles and photons is described by QED. This theory is based on the Abelian group $U(1)$, and its Lagrangian, $\mathcal{L}_{QED}$, is given by the following expression:

$$\mathcal{L}_{QED} = \bar{\psi} (i \not{D}_{QED} - m) \psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \quad (2.5)$$

where ψ denotes the charged fermion with mass m and charge Q , and F is the electric field tensor expressed by the photon vector field A , while $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. The Lagrangian in the case of a free particle is given by:

$$\mathcal{L} = \bar{\psi} \gamma^\mu \partial_\mu \psi - m \bar{\psi} \psi, \quad (2.6)$$

The electromagnetic interaction plays an important role at the microscopic level, having the function of forming atoms from the nucleus and electrons. Quantum field theory was originally used to describe the QED. It is considered the first of the forces described by it.

2.1.1.2 Quantum Chromodynamics

QCD mainly describes the interactions between color particles and it comes from the simple Lie group of 8 dimensions, $SU(3)$. There is some similarity between the QCD Lagrangian, $\mathcal{L}_{QCD}$ and the QED Lagrangian, which is an interaction term between a fermion field and a boson gauge field plus a field tensor term. However, with the exception that the gluon field tensor $G_{\mu\nu}^\alpha$ in QCD replaces the electric field tensor $F^{\mu\nu}$ in QED:

$$\mathcal{L}_{QCD} = \sum_n \bar{\psi}_n^b (i \not{D}_{QCD} - m_n) \psi_n^b - \frac{1}{4} G_{\mu\nu}^\alpha G^{\alpha\mu\nu}, \quad (2.7)$$

where b and α represent the three color charges index and the gluon index, respectively, while the covariant derivative $\not{D}_{QCD}$ and gluon tensor field are calculated using the following equations:

$$\not{D}_{QCD} = \gamma^\mu (\partial_\mu + i g_s T^\alpha A_\mu^\alpha), \quad (2.8)$$

μ and ν are Lorentz indices $(0, 1, 2, 3)$, γ^μ represents the Dirac matrices.

$$G_{\mu\nu}^\alpha = \partial_\mu A_\nu^\alpha - \partial_\nu A_\mu^\alpha - g_s f^{ijk} A_\mu^j A_\nu^k. \quad (2.9)$$

The strong nuclear force acting on fermions is described by QCD, mediated by gauge bosons called gluons. The gauge group $SU(3)_C$, where C denotes the color charge, can be used to derive the strong force. In this theory, gluons serve as force carriers, transmitting the strong force between quarks and other gluons.

One of the best-known properties of QCD is known as confinement which occurs at low energies. This is the reason why color-charged particles cannot be isolated or directly detected. In contrast to confinement, there is another property in the form of asymptotic freedom that occurs at high energies.

2.1.2 Gauge Bosons and Interactions

Gauge bosons describe the bosonic particles that act as force carriers in particle physics. These are the exact elementary particles for which gauge theory describes the interactions. By exerting forces on each other, these particles exchange gauge bosons. There are four types of gauge bosons corresponding to the three SM interactions, namely photons for the electromagnetic interaction, W and Z bosons for the weak interaction, and gluons for the strong interaction. All gauge bosons are considered vector bosons since they have a spin of 1. In contrast, the Higgs boson has a spin of 0 because it corresponds to a scalar field. Particles such as neutrinos interact only via weak interaction, while charged leptons and quarks undergo weak electromagnetic interaction and all three fundamental interactions, respectively. The SM contains all that is known about subatomic particles. The interactions of the particles are determined by the SM Lagrangian equation 2.10, which is shown below with its elements:

$$\mathcal{L}_{SM} = \mathcal{L}_G + \mathcal{L}_F + \mathcal{L}_Y + \mathcal{L}_H, \quad (2.10)$$

where $\mathcal{L}_G$ denotes the couplings to gauge bosons, $\mathcal{L}_F$ denotes the couplings to fermions, and $\mathcal{L}_Y$ and $\mathcal{L}_H$ denote the Yukawa couplings and the Higgs boson couplings, respectively. These components are defined as:

$$\mathcal{L}_G = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (2.11)$$

$$\mathcal{L}_F = i\bar{\psi} \not{D} \psi + h.c., \quad (2.12)$$

$$\mathcal{L}_Y = \bar{\psi} y_i \psi \phi + h.c., \quad (2.13)$$

$$\mathcal{L}_H = -|D_\mu \phi|^2 - V(\phi), \quad (2.14)$$

and they include kinetic terms of different fields. $\mathcal{L}_G$ and $\mathcal{L}_F$ include, in addition to the kinetic terms, the self-interactions of gauge bosons and the gauge interaction terms of fermions, respectively. While $\mathcal{L}_H$ contains the gauge boson interactions and the self-interaction of the Higgs boson, $\mathcal{L}_Y$ contains all interactions of the Higgs boson with leptons and quarks [28].

The three forces present in SM are described by the non-Abelian symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The strong force and QCD are described by $SU(3)_C$, while the weak force and the strong force, as well as electromagnetism and quantum electrodynamics, are described by $SU(2)_L$ and $U(1)_Y$, respectively. In addition, $SU(2)_L$ and $U(1)_Y$ can undergo electroweak unification to produce the electroweak sector explained by the Glashow-Weinberg-Salam theory [21, 22, 29]. The mathematical framework of the SM is formed by the conjunction of various symmetries or interactions. Therefore, the SM is considered a quantum field theory based on $SU(3)_L \times SU(2)_L \times U(1)_Y$ gauge symmetry.

2.1.3 Elementary Fermions

Fermions are known to be the building blocks of matter. The SM consists of 12 of these fermions, which are divided into six quarks and six leptons. For each matter particle, there is an antimatter particle that has the same mass but is opposite in other forms. These particles interact via gauge bosons, which is also known as the exchange of force-carrying particles. The gauge bosons have an integer spin, while the fermions have a half spin. The six leptons and quarks in each generation are grouped into three generations (see Figure 2.1). The quark generations consist of up and down, charm and strange, and top and bottom, as shown in Figure 2.1. Electron and electron neutrino, muon and muon neutrino, and tau and tau neutrino form the leptons. Electron, muon, and tau carry -1 electric charge. The interaction between them and neutrinos is weak because neutrinos carry no charge. Quarks, unlike leptons, can interact with all forces and have a charge of $+2/3$ and $-1/3$. The up quark has a charge of $+2/3$, while the down quark has a charge of $-1/3$. The quarks are divided into three generations, each differing only in mass. They also interact with each other via a strong force, and each of them has a color charge quantum number. Quarks and leptons are the two types of fermions, with quarks combining to form protons and neutrons. One down quark and two up quarks form the proton, which is held together by a strong nuclear force. Leptons, on the other hand, are unaffected by this force.

The SM describes matter and antimatter that have opposite charges compared to their counterparts. Fermions have an antimatter pair or counterpart and these are treated the same as their matter equivalent by SM. The theory explains how quarks and leptons are part of all known matter. Protons make up protons and neutrons, while leptons make up electrons. In addition, this theory also explains how force-carrying particles affect the quarks and leptons. The force-carrying particles belong to a group of bosons.

Quarks come in three colors: red, green, and blue, and they have the ability to stick to other quarks because they have the color charge. Gluons are the forces that hold these colored particles together. Since gluons are also colored, they also bind themselves to each other. QCD is the name for this interaction, which involves the exchange of gluons.

2.2 The BEH Mechanism and The Higgs Boson

The BEH (Brout-Englert-Higgs) mechanism explains how particles get their mass. The Higgs boson, unlike the gauge boson, which is a type of force carrier, is an elementary particle created by quantum excitation of the Higgs field. As opposed to gauge bosons that have a spin of 1, the Higgs bosons have a spin of 0. Unlike the gauge boson, which is a vector boson, the Higgs boson is a scalar boson. It is a massive neutral boson that interacts with other particles to give them mass but does not carry a force.

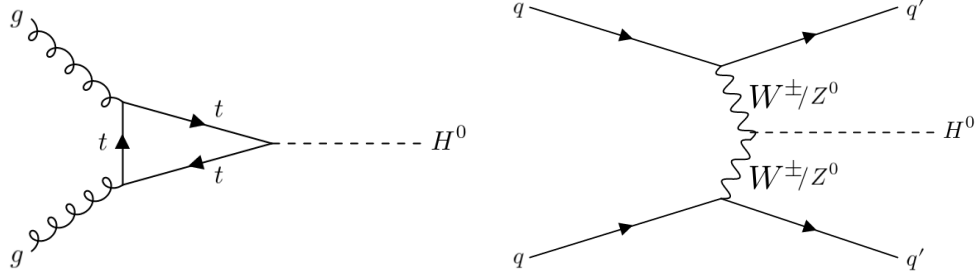
The BEH mechanism allows quarks and leptons to determine their masses. It was discovered in 2012 by the ATLAS and CMS teams [18, 30] and completed the SM puzzle. This discovery was of great importance to particle physics, as it proved the effectiveness of the SM [21–23] and confirmed its mass of 124.97 ± 0.24 GeV [31]. The measurement

of its mass was performed in the Run 1 and Run 2 datasets in the $H \rightarrow ZZ \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels. After the first measurement of the Higgs boson, both ATLAS and the CMS experiments conducted extensive research. This research focused on some of the properties of this particle, with the goal of confirming that they are consistent with those predicted by the SM. It is considered to be a massive boson with no electric charge. Furthermore, since this boson has no color charge and is unstable, it can easily decay into other particles. The boson takes its name from physicist Peter Higgs. The primary production mechanisms for the SM Higgs boson include gluon fusion (ggF), vector boson fusion (VBF), and Higgs-Strahlung with an associated vector boson [32], as shown in Figure 2.2. The ggF mechanism, which serves as the dominant production, exhibits a cross-section of approximately 50.35 pb for a Higgs boson mass of 125 GeV at $\sqrt{s} = 14$ TeV [33]. The VBF process is associated with a cross-section of around 4.172 pb. The Higgs-Strahlung mechanism, on the other hand, has a much smaller effective cross-section than other mechanisms, and one of its advantages is that it has an experimental signature of a massive vector boson or top quark. The cross-section of the associated production with a Z and W boson is about 0.8830 pb and 1.504 pb, respectively. Finally, the $t\bar{t}H$ Higgs production exhibits a cross-section of 0.6113 pb. The SM, on the other hand, can predict all the properties of the Higgs boson as well as the quantum numbers, except for the Higgs boson mass.

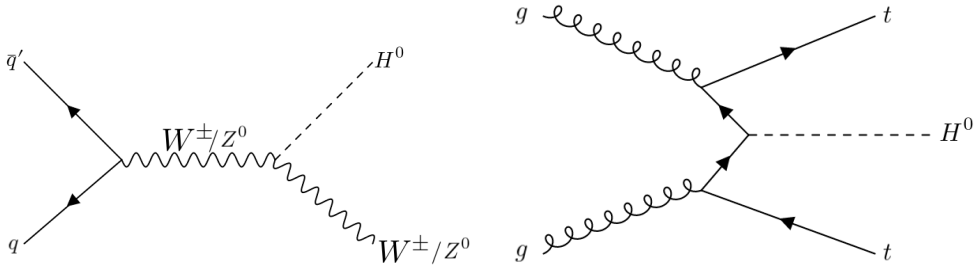
2.3 Beyond the Standard Model

Although SM is recognized as the most successful theory for accurately predicting the quark-lepton interaction in particle physics, it also has its limitations. Because of these limitations, physicists are searching for a more complete theory. Among the limitations of the current SM is that it cannot explain why there are three generations of quarks and leptons and why neutrinos are massless or have a very small mass. Therefore, it is unable to predict the strength of certain interactions and the masses of quarks and leptons. It also cannot explain the existence of dark matter and dark energy in the universe or the exclusion of the gravitational force from other fundamental forces. As physicists move beyond this model and make more accurate measurements, they hope to find a more complete theory.

Several theoretical solutions have been proposed to overcome the limitations of the SM. The supersymmetry (SUSY) [34] and the two Higgs doublet models (2HDM) [35, 36] are examples. The SUSY model ensures that each particle has a superpartner that has a spin that differs by half an integer. This adds a new symmetry between fermions and bosons by changing the spin of the particle by $\frac{1}{2}$. This transformation turns fermions into bosons and vice versa. S-fermions are fermion superpartners, while boson superpartners are bosinos. The overall superpartners of the SM particles, on the other hand, are called s-particles. It is known that the Higgs boson is the source of mass for quarks, leptons, and the W and Z bosons [37]. For the neutrinos, however, this is different or rather not known.



(a) Gluon-gluon fusion where Higgs boson is produced through the interaction of two gluons via top quarks. (b) Vector boson fusion, where the Higgs boson together with two final state quarks are produced by two vector bosons



(c) Higgs-Strahlung, where a W or Z boson is produced in association with a Higgs boson (d) $t\bar{t}H$ production, where a top quark pair is produced in association with a Higgs boson

Figure 2.2: The primary production mechanisms for the SM Higgs boson at high energy pp collisions.

2.3.1 The Madala Hypothesis

The discovery of the Higgs boson is consistent with the SM hypothesis. There have also been reports of a number of excesses in the Run 1 ATLAS and CMS data that are being studied by physicists. These excesses may indicate the existence of a second, heavier scalar particle. One of the more recently formulated hypotheses is the Madala boson (described in detail in [7, 38]), which was developed in response to the anomalies in the Run 1 ATLAS and CMS data. The term Madala means “an old man” in isiZulu, one of the official languages of South Africa. This was formulated by the Wits Institute for Collider Particle Physics (ICPP) in collaboration with researchers at the Harish-Chandra Research Institute (HRI) in India.

This theory introduces the presence of two scalars, a hypothetical heavy scalar particle H and a Higgs-like scalar S . The heavy scalar is known as the Madala boson. Since it yields two new heavy scalars heavier than the SM Higgs boson, this hypothesis serves as an extension of the SM. Therefore, it was used to explain discrepancies in the LHC Run 1 data. It was assumed to have Higgs-like couplings with the SM particles. It is also thought to be a resonant di-Higgs production source. This appeared to resemble some of the features associated with the production of the Higgs boson, which showed

Result	Publication	
Differential Higgs boson p_T spectra	ATLAS	Fiducial cross section measurements on $h \rightarrow \gamma\gamma$ [7] and $h \rightarrow ZZ^* \rightarrow 4l$ [40]
	CMS	Fiducial cross section measurements on $h \rightarrow \gamma\gamma$ [41] and $h \rightarrow ZZ^* \rightarrow 4l$ [42]
Di-Higgs boson resonance searches	ATLAS	Limits on $H \rightarrow hh \rightarrow b\bar{b}\tau\tau, \gamma\gamma WW^*, \gamma\gamma b\bar{b}$ and $b\bar{b}b\bar{b}$ [43]
	CMS	Limits on $H \rightarrow hh \rightarrow \gamma\gamma b\bar{b}$ [44], $b\bar{b}\tau\tau$ [45] and multilepton [46]
Top associated Higgs boson production	ATLAS	Limits on $h \rightarrow \gamma\gamma$ [47], Measurements on multi-lepton decay channels [48] and $h \rightarrow b\bar{b}$ [49]
	CMS	Limits on $h \rightarrow \gamma\gamma$ and $b \rightarrow b\bar{b}$ [50]
$H \rightarrow VV$ decays	ATLAS	Limits on $H \rightarrow WW$ [51] and $H \rightarrow ZZ$ [52]
	CMS	Limits on $H \rightarrow WW$ and $H \rightarrow ZZ$ [53]

Table 2.1: ATLAS and CMS results utilized to constrain the relevant parameters of the Madala hypothesis [7].

some minor deviations from the SM predictions. The deviations were found in searches for VV and hh resonances around a mass of 275 GeV [7, 39]. The Madala hypothesis predicts that the Higgs-like heavy scalar boson H originates from a ggF production mode and decays into the Higgs boson and some unidentified particles X , $H \rightarrow h + X$. This decay has somehow boosted the Higgs boson to higher values of p_T which helps explain the excess in the Higgs p_T spectra. Subsequently, this has motivated and encouraged speculation about BSM physics to some extent.

The predicted heavy scalar is believed to have a mass of 272_{-9}^{+12} GeV and has been quantified as a three-sigma effect [7]. It is worth noting that this hypothesis was based on the ATLAS and CMS measurements presented in Table 2.1. These results come from the spectra of the Higgs bosons $p_{T,h}$, the search for di-Higgs bosons, the production of top-associated Higgs, and the VV decays. However, these discrepancies, explained by each study individually, are not significant enough to be considered as potential for new physics. Moreover, they can be jointly explained by the formulated hypothesis [7, 39].

The formulated Madala hypothesis was successfully used to explain the discrepancies in the Run 1 and Run 2 data at the LHC [54]. The heavy scalar H shows a strong interaction with the SM Higgs boson and the additional Higgs-like scalar singlet S [38]. When applied to various experimental results from Run 1, the hypothesis has constrained the mass of H to $m_H = 272_{-9}^{+12}$ GeV [7], while the mass of S has not yet been constrained, but is believed to be in the range of $130 \leq m_S \leq 200$ GeV [54]. This can be attributed to the fact that this range has a rich combination of multiple lepton production that can be probed, and that the S additionally decays into pairs of massive gauge bosons. Studies carried out on this formulated hypothesis have been mainly looking at the mechanism of gluon fusion (ggF) to produce H , which can also be produced in association with the top quarks ($t\bar{t}H$).

The Lagrangian for the simplified model of H and S is given by:

$$\mathcal{L} = \mathcal{L}_{SM} + \mathcal{L}_{BSM}, \quad (2.15)$$

where $\mathcal{L}_{BSM}$ includes the new interactions and states. These new interactions are assumed to originate from effective couplings, with the Higgs boson having normal SM interactions with the fermions and gauge bosons. Thus, the proposed BSM Lagrangian

with the new scalars includes:

$$\mathcal{L}_{\text{BSM}} \supset \mathcal{L}_H + \mathcal{L}_Y + \mathcal{L}_T + \mathcal{L}_Q, \quad (2.16)$$

Where $\mathcal{L}_Y$, $\mathcal{L}_T$ and $\mathcal{L}_Q$ terms represent the Yukawa, tri-linear, and quartic interactions, respectively, defined as follows:

$$\mathcal{L}_H = -\frac{1}{4}\beta_g\kappa_{hgg}^{\text{SM}}G_{\mu\nu}G^{\mu\nu}H + \beta_V\kappa_h^{\text{SM}}V_\mu V^\mu H, \quad (2.17)$$

$$\mathcal{L}_Y = -\frac{1}{\sqrt{2}}[y_{ttH}\bar{t}tH + y_{bbH}\bar{b}bH], \quad (2.18)$$

$$\mathcal{L}_T = -\frac{1}{2}v[\lambda_{Hhh}Hhh + \lambda_{h\chi\chi}h\chi\chi + \lambda_{H\chi\chi}H\chi\chi], \quad (2.19)$$

$$\begin{aligned} \mathcal{L}_Q = & -\frac{1}{2}\lambda_{Hh\chi\chi}Hh\chi\chi - \frac{1}{4}\lambda_{HHhh}HHhh \\ & -\frac{1}{4}\lambda_{hh\chi\chi}hh\chi\chi - \frac{1}{4}\lambda_{HH\chi\chi}HH\chi\chi, \end{aligned} \quad (2.20)$$

H and χ represent the heavier scalar and the dark matter candidate, respectively. While t and b represent the fields of the top quark and the bottom quark, $\bar{t}$ and $\bar{b}$ represent the Dirac conjugates of the top quark and the bottom quark, respectively.

The simplified Lagrangian model for S is given by:

$$\mathcal{L}_S = \mathcal{L}_K + \mathcal{L}_{SVV} + \mathcal{L}_{Sf\bar{f}} + \mathcal{L}_{hHS} + \mathcal{L}_{S\chi}, \quad (2.21)$$

where

$$\mathcal{L}_K = -\frac{1}{2}\partial_\mu S\partial^\mu S - \frac{1}{2}m_S^2 SS, \quad (2.22)$$

$$\begin{aligned} \mathcal{L}_{SVV} = & \frac{1}{4}\kappa_{Sgg}\frac{\alpha_S}{12\pi v}SG^{a\mu\nu}G_{\mu\nu}^a + \frac{1}{4}\kappa_{S\gamma\gamma}\frac{\alpha}{\pi v}SF^{\mu\nu}F_{\mu\nu} \\ & + \frac{1}{4}\kappa_{SZZ}\frac{\alpha}{\pi v}SZ^{\mu\nu}Z_{\mu\nu} + \frac{1}{4}\kappa_{SZ\gamma}\frac{\alpha}{\pi v}SZ^{\mu\nu}F_{\mu\nu} \\ & + \frac{1}{4}\kappa_{SWW}\frac{2\alpha}{\pi S_w^2 v}SW^{+\mu\nu}W_{\mu\nu}^-, \end{aligned} \quad (2.23)$$

$$\mathcal{L}_{Sf\bar{f}} = -\sum_f K_{Sf}\frac{m_f}{v}S\bar{f}f, \quad (2.24)$$

$$\mathcal{L}_{HhS} = -\frac{1}{2}v[\lambda_{hhS}hhS + \lambda_{hSS}hSS + \lambda_{HHS}HHS + \lambda_{HSS}HSS + \lambda_{HhS}HhS], \quad (2.25)$$

$$\mathcal{L}_{S\chi} = -\frac{1}{2}v\lambda_{S\chi\chi}S\chi\chi - \frac{1}{2}\lambda_{SS\chi\chi}SS\chi\chi, \quad (2.26)$$

where V represents the gluons, photons, and weak vector bosons. H , χ , and v represent the heavy scalar, the dark matter candidate, and the vacuum expectation value ($v = v_{ev} = 246$ GeV), respectively. Moreover, the term in Equation 2.17 represents the

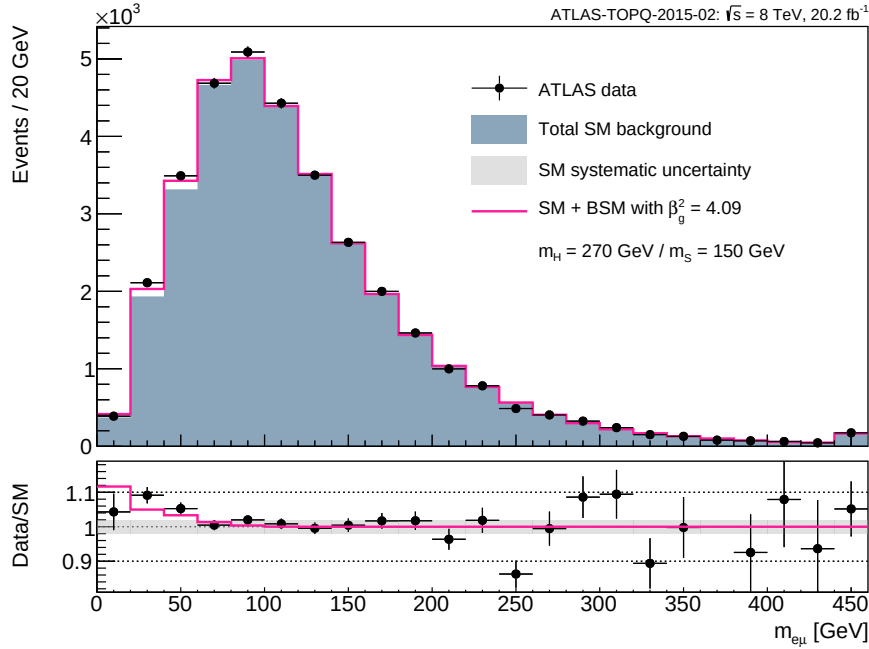


Figure 2.3: Dilepton invariant mass distribution of ATLAS data (dots) and SM background (red line) compared with the BSM signal (blue line) for events requiring a pair of oppositely charged leptons of a different flavor ($e^\pm\mu^\pm$) with at least one b -tagged jet. [56].

effective coupling of gluons to H , where κ_{hgg}^{SM} denotes the SM coupling defined as $\alpha_s/\pi v$. In addition, β_g serves as a regulating parameter for the production rate of $gg \rightarrow H$. This model is discussed in detail in [55]. The absence of a specific supersymmetric Lagrangian for the Higgs field can be attributed to the fundamental principles and mathematical formalism of supersymmetry. In supersymmetric theories, the Higgs field is part of a larger framework that incorporates a symmetry between fermions and bosons through the introduction of superpartners. These superpartners have different quantum statistics compared to their corresponding particles. Therefore, the treatment of the Higgs field within supersymmetry is not isolated, but rather integrated into the broader supersymmetric structure.

2.3.1.1 Testing the hypothesis in the multi-lepton final states

Some of the inconsistencies in the data can be explained by the Madala hypothesis, in particular the multi-lepton final states. The invariant mass spectra of the multiple leptons of ATLAS and CMS can be studied in the context of the new possibilities using the Madala hypothesis. Excess has been predicted for events with two leptons plus jets and E_{miss}^T [7, 38]. A study based on the decay of S into di-lepton final states was performed to understand this excess [4]. In this study, the BSM di-lepton signal was compared to both the ATLAS and CMS data, with the optimal performance achieved by changing the BSM standardization and fitting the SM +BSM approximation to the

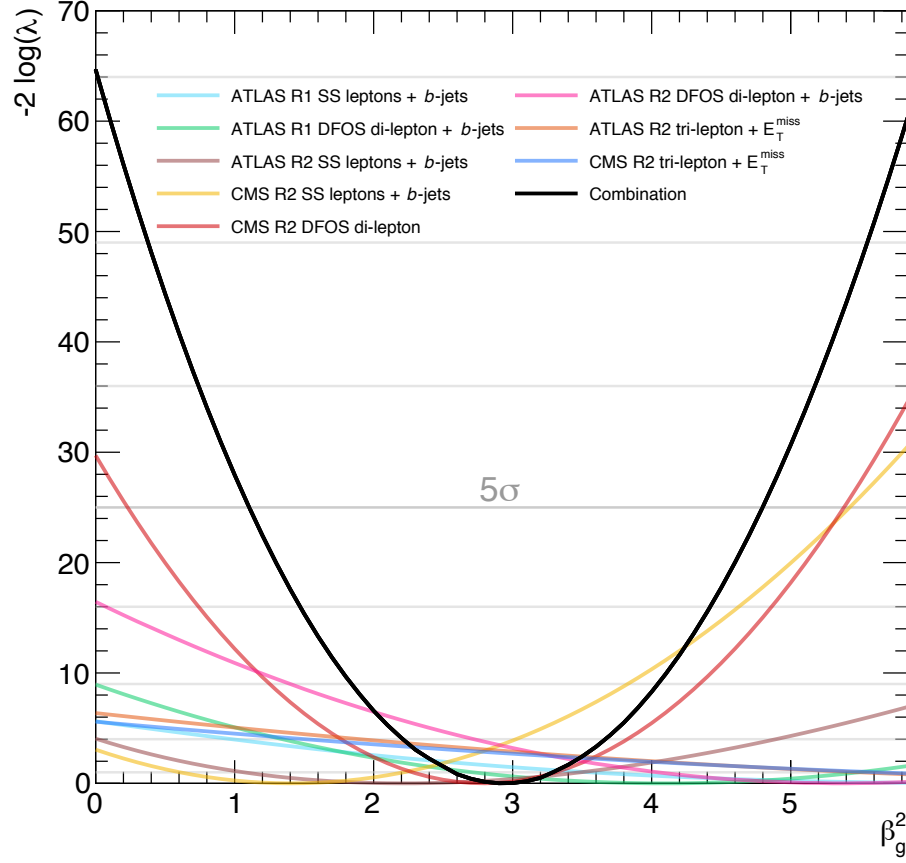


Figure 2.4: The (negative 2 times logarithm of the) profile likelihood ratios for each of the individual fit outcomes, overlaid on that of their combination [6].

results based on a series of measurements.

Selection	Best-fit β_g^2	Significance
ATLAS Run 1 SS $\ell\ell$ and $\ell\ell\ell + b$ -jets	6.51 ± 2.99	2.37σ
ATLAS Run 1 OS $e \mu + b$ -jets	4.09 ± 1.37	2.99σ
ATLAS Run 2 SS leptons $+b$ -jets	2.22 ± 1.19	2.01σ
CMS Run 2 SS leptons $+b$ -jets	1.41 ± 0.80	1.75σ
CMS Run 2 OS $e \mu$	2.79 ± 0.52	5.45σ
ATLAS Run 2 SS $\ell\ell$ and $\ell\ell\ell + b$ -jets	5.42 ± 1.28	4.06σ
ATLAS Run 2 tri-lepton $+E_{miss}^T$	9.05 ± 3.35	2.25σ
CMS Run 2 tri-lepton $+E_{miss}^T$	9.70 ± 3.88	2.36σ
Combination	2.92 ± 0.35	8.04σ

Table 2.2: A summarized results of the SM+BSM fit and their combination [6].

A further study of this can be found in Ref. [6] to investigate the same basic model with Higgs-like H and S and only one parameter to understand the anomalies in multi-lepton final states and test the Madala hypothesis, setting the masses to $m_H = 270$ GeV and $m_S = 160$ GeV, respectively. The goal is to determine how consistent these variances are. Figure 2.3, comparing the data from ATLAS, the background from SM, and the BSM signal after selecting events with two oppositely charged leptons of different flavors and at least one b -tagged jet shows the invariant mass distribution of the dileptons [56]. In the region below 100 GeV, the invariant mass distribution detects inconsistencies that the $H \rightarrow Sh$ signal can explain. All measures are fitted simultaneously, and the profile-likelihood ratio is shown in Table 2.2 and Figure 2.4 for each outcome and its combination [6]. The best fit is $\beta_g^2 = 2.92 \pm 0.35$, which differs from SM by 8.04σ . This shows that the existing instruments are not able to explain multi-lepton data and that our simpler model can capture the effect with only one parameter. Moreover, the idea of incorporating the Madala hypothesis into a 2HDM+S model is now well-positioned to be further explored by experiments in multi-lepton final states.

Chapter 3

The Large Hadron Collider and the ATLAS Detector

In this chapter, we give a brief description of the LHC in Section 3.1, where Sections 3.2 and 3.3 focus on the ATLAS detector and trigger system, respectively, while Section 3.4 presents the event reconstruction.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest circular superconducting machine in the world, with a circumference of 27 kilometers and a depth of 100 meters [57]. It is operated by the European Organization for Nuclear Research (CERN) [58] and is located on the French-Swiss border. This machine generates a huge amount of data from proton beam collisions with a centre-of-mass energy of up to $\sqrt{s} = 14$ TeV. The protons are accelerated to energies of up to 7 TeV, and the collisions at the LHC produce particles that physicists are exploring in search of new physics beyond the Standard Model (BSM) and to test the SM at high energies. To increase the probability of BSM events occurring, particles must be accelerated to extremely high energy and luminosity before they collide.

As shown in Figure 3.1, this machine has four collision sites where protons collide with each other. After being accelerated to greater energy by the LHC's succession machines [59], they are injected into the LHC. Protons from a hydrogen source are accelerated in several stages. The first of these is the linear accelerator Linac 2, which accelerates protons to an energy of 50 MeV. This is followed by the Proton Synchrotron Booster (PSB), which accelerates protons to an energy of 1.4 GeV. Proton Synchrotron (PS) and Super Proton Synchrotron (SPS) are the last two acceleration processes that accelerate protons to energies of 25 GeV and 450 GeV, respectively, before they are injected in bunches, in opposite directions into the LHC ring. As they travel in opposite directions and collide at specific collision points, at their maximum center of mass energy [60]. Particle detectors surround these collision points and capture the particles that emerge from the collision to reconstruct the physics of interest.

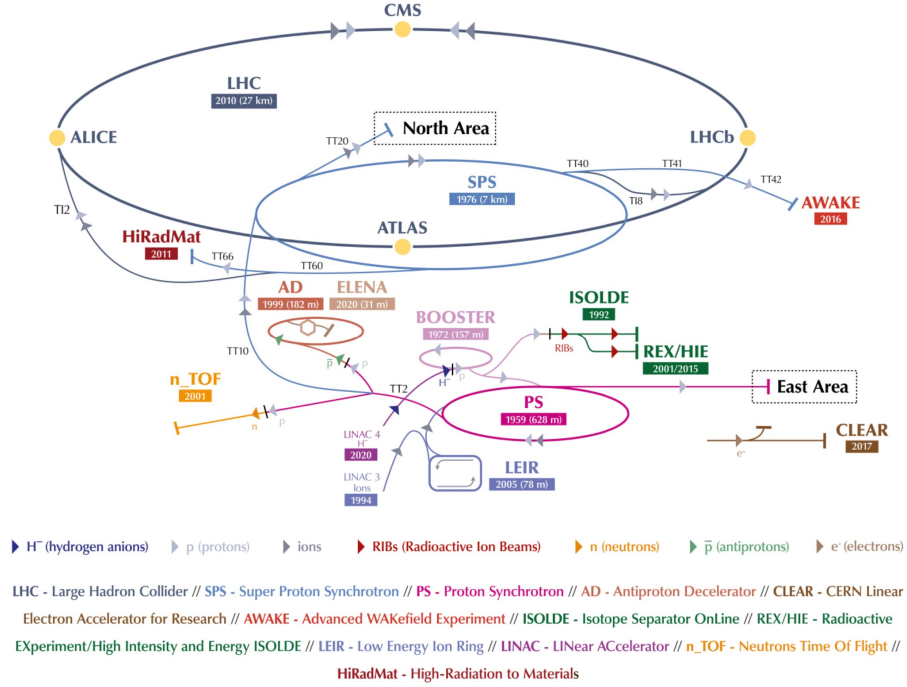


Figure 3.1: An illustration of the CERN accelerator complex [61].

ATLAS (A Toroidal LHC Apparatus) [62], CMS (Compact Muon Solenoid) [63], LHCb (LHC Beauty) [64], and ALICE (A Large Ion Collider Experiment) [65] are the four main detectors at the LHC, each placed at one of the four interaction points. ATLAS and CMS are both interested in the discovery of new heavy particles, while ALICE and LHCb study the properties of the quark-gluon plasma in heavy-ion collisions and the antisymmetries between matter and antimatter, respectively. The luminosity phenomenon describes a quantity that measures the number of collisions produced or the probability of a quantum mechanical event by the LHC. This yields the proportionality of the event rate and the effective cross-section:

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma, \quad (3.1)$$

where $\frac{dN}{dt}$ represents the number of events per second, σ and $\mathcal{L}$ represent the cross-section of protons collisions and luminosity [67], respectively. However, this describes an instantaneous luminosity, the actual number of events is obtained by integrating the equation 3.1 over the time period in which collisions occur, *i.e.* integrated luminosity. This actual number is given by the following equation:

$$N = \int \mathcal{L} dt \cdot \sigma, \quad (3.2)$$

where $\mathcal{L}$ is defined as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} \times F, \quad (3.3)$$

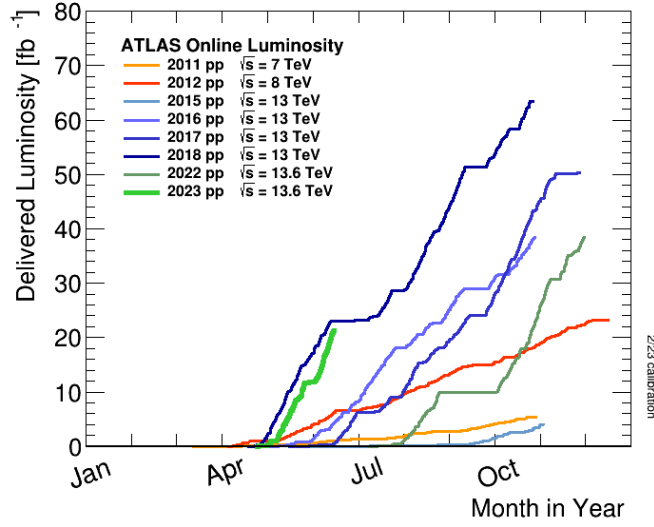


Figure 3.2: The recorded ATLAS integrated luminosity in different years [66].

and:

- N_b is the number of particles per bunch,
- n_b is the number of bunches per beam,
- f_{rev} is the revolution frequency,
- γ_r is the relativistic gamma factor,
- ϵ_n is the normalized transverse beam emittance,
- β^* is the collision point beta function, defined by:

$$\beta^* = \frac{\alpha^2 \sigma_r}{\epsilon_n}, \quad (3.4)$$

- and F is the geometric luminosity factor, defined as:

$$F = \frac{1}{\sqrt{1 + (\frac{\theta_c \sigma_z}{2\sigma^*})^2}}, \quad (3.5)$$

where θ_c is the crossing angle of the beams, σ_z and σ^* represent the root-mean-square bunch length and root-mean-square transverse beam size at the interaction point, respectively.

To date, the LHC machine has completed two runs. Run 1 took place in 2010-2011 and 2012, with $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV as the centre-of-mass energy, respectively. After a two-year maintenance break to improve the accelerators and detectors, Run 2 ran from 2015 to 2018, with the centre-of-mass energy for this run raised up to 13

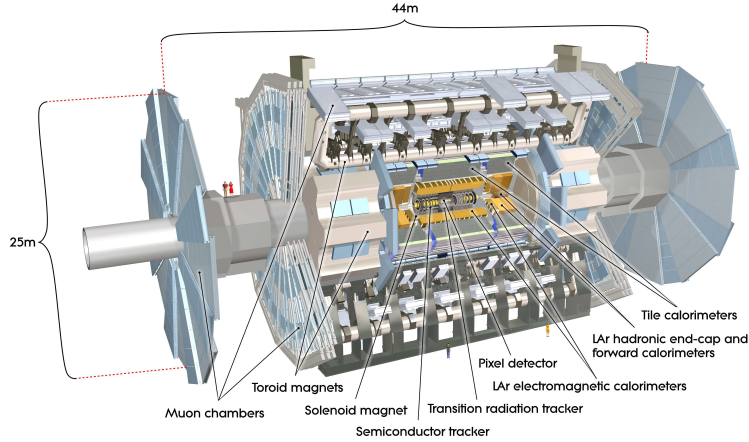


Figure 3.3: The schematic illustration of the ATLAS detector [62].

TeV. Run 3 started in 2022 and will run until 2025 at 13.6 TeV. Figure 3.1 shows the integrated luminosity over time. The sum of the luminosity per year determines the total luminosity for a given run. For example, the total luminosity for Run 2 is calculated by adding 3.2 fb^{-1} for 2015, 32.9 fb^{-1} for 2016, and 43.7 fb^{-1} and 60.1 fb^{-1} for 2017 and 2018, respectively, for a total of 140 fb^{-1} . The unit is given in inverse cross section because the number of events is adimensional. The luminosity for Run 3 is 56 fb^{-1} .

3.2 The ATLAS Detector

The ATLAS detector, shown in Figure 3.3, is the largest LHC particle detector and acts as a camera that records and observes data from collisions. It is also used to measure common SM physics and search for new physics BSM. This structure, which is 25 meters high, 44 meters long, and weighs over 7000 tons, is located along the LHC ring. The detector is divided into three parts, each with its own particle identification system. The Inner Detector, the calorimeter, and the muon spectrometer are the subsections surrounding the proton-proton (pp) interaction point and they are also discussed in detail in the following subsections. The detector was designed with the primary goal of studying collisions. The Inner Detector, composed of the Pixel detector, the Semiconductor Tracker, and the Transition Radiation Tracker, is responsible for monitoring charged particles and sampling both the electromagnetic and hadronic calorimeters for energy measurements. A muon spectrometer, on the other hand, tracks muons. Figure 3.3 shows a schematic diagram of this detector, detailing the various sections of the ATLAS in the LHC. The muon spectrometer makes use of three air-core toroid magnets of 8 Tesla with a thin superconducting solenoid of 2 Tesla covering the Inner Detector. The solenoid is responsible for generating a magnetic field that can be used to measure the momentum of charged particles for identification.

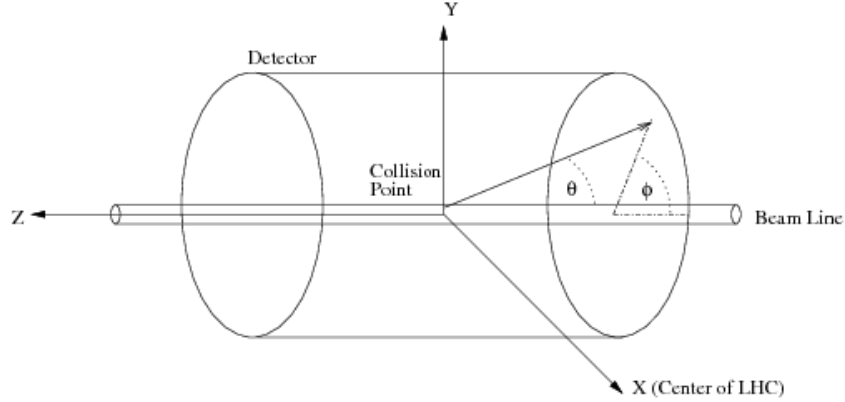


Figure 3.4: The ATLAS coordinate system showing the collision point in the Cartesian plane xyz with polar and azimuthal angles [62].

3.2.1 ATLAS Coordinate System

The ATLAS detector uses a right-handed (x, y, z) coordinate system. The positive x -axis and the positive y -axis each refer to the center of the LHC and upward toward the sky, while the z points along the beam axis. The xy -plane is called the transverse plane, and transverse components are vector components in this plane. The transverse momentum p_T is the component of the momentum that is perpendicular to the beam line and is measured by:

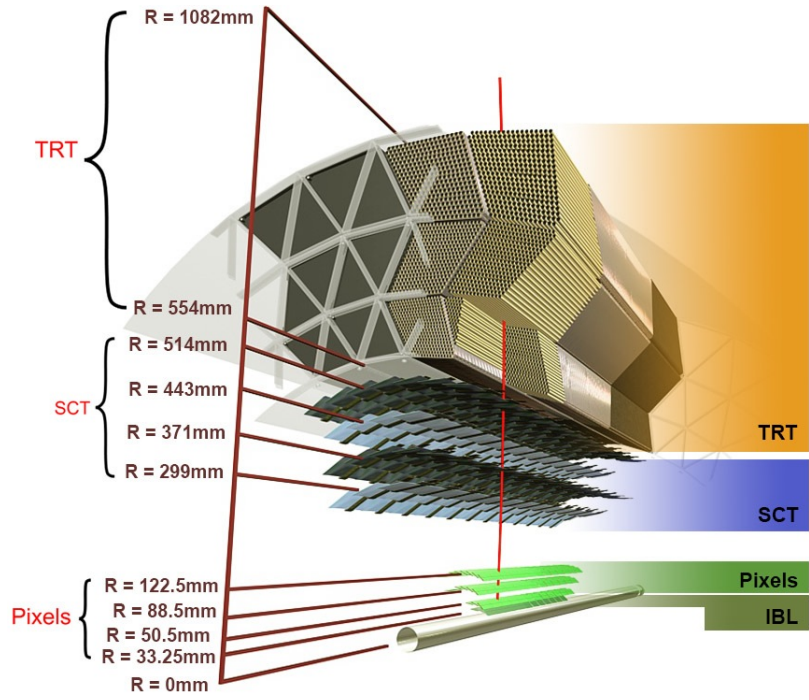
$$p_T = \sqrt{p_x^2 + p_y^2}. \quad (3.6)$$

Figure 3.4 shows the coordinate system, which includes both the Cartesian and polar planes. The cylindrical shape of the detector justifies the use of polar (r, ϕ, z) or spherical (ρ, ϕ, θ) coordinates, where $\phi = \arctan(y/x)$ is the azimuthal angle around the z -axis (beam axis), the polar angle $\theta = \arctan(r/z)$ measures the rotation from the beam axis and $r = \sqrt{x^2 + y^2}$ is the distance from the beam pipe. The polar angle θ is also used to define the pseudorapidity $\eta = -\ln \tan(\theta/2)$, which is the rapidity limit for highly relativistic particles. However, the rapidity is measured by the following equation:

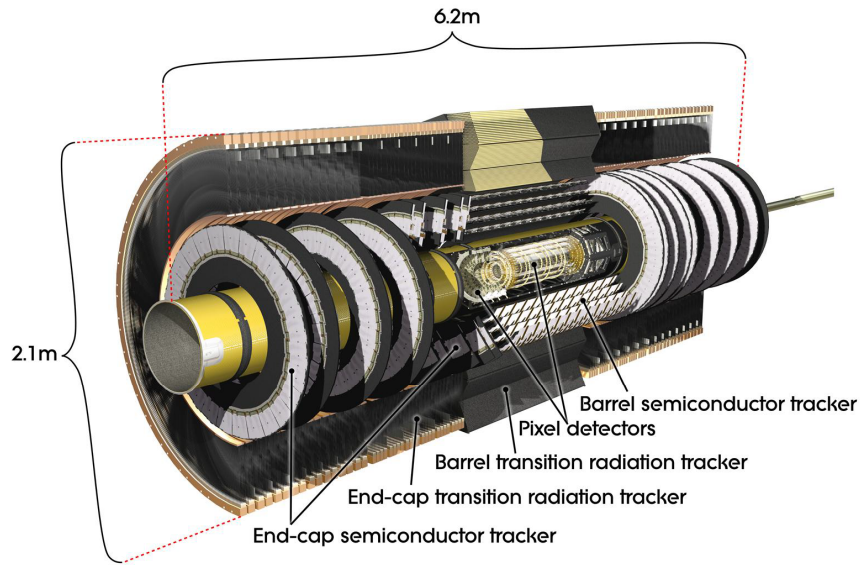
$$Y = \frac{1}{2} \ln \frac{E + P_z}{E - P_z}, \quad (3.7)$$

where E and p_z are the energy and momentum of the particle in the z -direction, respectively. The ATLAS detector uses the $\eta - \phi$ coordinate space to describe position and direction, where particles existing perpendicular and parallel to the beam axis have η of 0 and $\pm\infty$, respectively. Also, the distance between particles is calculated using the following expression:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \quad (3.8)$$



(a)



(b)

Figure 3.5: (a) The cutway view of the ATLAS Inner Detector system [68], (b) The Inner Detector of ATLAS [69].

3.2.2 Inner Detector

The Inner Detector (ID), shown in Figure 3.5(a), is a layer consisting of four primary subdetectors occupying about 1 m of the innermost part of the ATLAS detector. The ID is 7 meters long and 2.3 meters wide and has a pseudorapidity of $\eta \leq 2.5$ [62]. It is also closest to the interaction point and the beam pipe. It is thus responsible for the detection of charged particles and the reconstruction of their paths and is embedded in a 2 Tesla solenoid that ensures the precise measurement of the momentum of the charged particles. In addition to accurate measurement, the solenoid is also responsible for the deflection and bending of the charged particles. The Pixel detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT) are the three subdetectors that make up the ID, and together they provide millions of readout channels. Particles go outward from the collision point through these subdetectors in the following order: Pixel detector, SCT, and TRT.

3.2.2.1 Pixel Detector

The Pixel detector consists of three barrel layers and three endcap disks that provide coverage of $|\eta| < 2.5$, providing high granularity and resolution on the order of microns. The three-barrel layers are 50.5 mm, 88.5 mm, and 122.5 mm away from the centre of the beam pipe. This is the most populated layer of the ATLAS detector, with approximately 80 million readout channels and 1744 modules of 46080 pixels each, each pixel segmented in the $R-\phi$ plane and along z and R , with resolutions of $10 \mu\text{m}$ and $115 \mu\text{m}$, respectively. During the first long upgrade (2013-2014), just before the start of Run 2, an additional fourth layer was installed in the form of an insertable B-layer (IBL). This was installed between the existing Pixel detector and a smaller radius beam pipe to handle high radiation and occupancy. This layer has a length of 1.4 meters and a diameter of 0.5 meters. It is located at a distance of $R = 33.23 \text{ mm}$ from the z -axis [70].

3.2.2.2 Semiconductor Tracker

The SCT consists of four-barrel layers, each with 4088 two-sided silicon microstrip modules [71]. SCT uses the same techniques as the Pixel detector, except that it consists of long strips of tracking silicon boards. There are about 6.3 million readout strips in these units. Its main function is to measure momentum and it covers a pseudorapidity range of $|\eta| < 2.5$. The SCT is 300 mm from the beam pipe and covers the Pixel detector. This is done to improve accuracy in tracking the curvature of charged particles at greater distances from the Pixel detector, where the particles would spread further. Its modules are also segmented perpendicular and parallel to the strips in the $R-\phi$ plane and to the beam in z and R , respectively. The former has an accuracy of $17 \mu\text{m}$, while the latter has $580 \mu\text{m}$ [72].

3.2.2.3 Transition Radiation Tracker

The TRT is the last and outermost component of the ID and follows the SCT. Unlike SCT and ID, TRT measures only azimuthal information with a resolution of $130 \mu\text{m}$ in ϕ and has a pseudorapidity range of $|\eta| < 2.0$ [62]. Unlike silicon-based detectors and

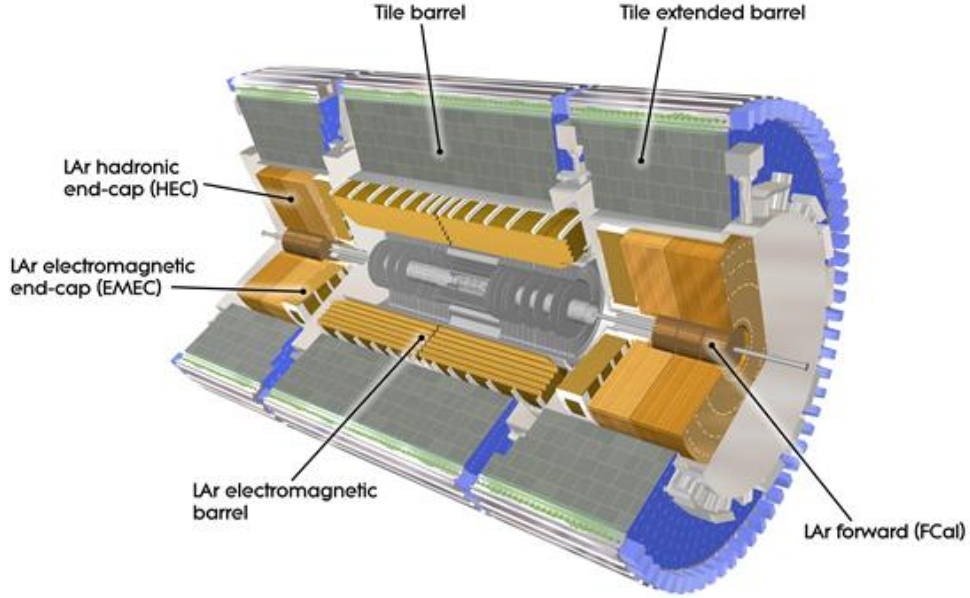


Figure 3.6: The cutway view of the ATLAS calorimeter system [62].

SCT, it consists of multiple gas-filled straw tube detectors, each 4 mm in diameter. In the barrel and end-cap sections, these straws are organized in modules of 52544 and 122880, respectively, and filled with a gas mixture of Xe (70%), CO₂ (27%), and O₂ (3%), which is ionized by the charged particles as they pass through. TRT is the largest subdetector in the Inner Detector [73, 74], and it detects the transition radiation used to identify the particles. However, it has a low intrinsic resolution, allowing measurements only in the $R - \phi$ plane. This results in a precision of $130 \mu m$ per straw.

3.2.3 Calorimeters

The calorimeter system shown in Figure 3.6 is an energy detector that works by stopping particles and absorbing their energy. This system, which consists of nested inner and outer layers called the electromagnetic calorimeter (ECal) and hadronic calorimeter (HCal), is located behind the Inner Detector. It is also located outside the solenoidal magnet [62]. The ATLAS calorimeter is used to determine the energy of particles and jets and to accurately measure the missing transverse energy E_T^{miss} and covers the range $|\eta| < 4.9$. When high-energy particles collide with dense absorbing material, particle showers are formed. This produces low-energy particles that interact with the sample medium. The result is a signal in the form of photons or an electric current proportional to the output energy. The ECal has the task of separating electromagnetically interacting particles with low mass, such as electrons and photons, from particles with high mass, which are dominated by strong interactions. Protons, neutrons, pions, and other hadronic particles are examples of these particles. Each section of the calorimeter can be divided into active and inactive zones, also called absorbing mediums. The absorb-

ing medium is responsible for slowing down the incoming particles and promoting the showers, while the active region captures the energy of the resulting showers.

3.2.3.1 Electromagnetic Calorimeter

Like the sampling calorimeter, the ECal uses liquid Argon as the active medium and lead as the absorption medium [75]. It detects showers and measures the energy of electrons and photons over an azimuthal range of up to $|\eta| < 3.2$ [75]. This is divided into two parts: a barrel with a range of $|\eta| < 1.475$ and two end-caps with a range of $1.375 < |\eta| < 3.2$. ECal is the first layer of the calorimeter system that closely monitors the energy of photons, electrons, jets, and E_T^{miss} . The barrel has three sample layers, the first two of which have a thickness of 4 and 16 radiation lengths, respectively. This ensures that only a small portion of the electromagnetic shower is deposited on this layer, while the majority is deposited on the second layer. In addition, the third layer is only responsible for distinguishing electromagnetic from hadronic showers and is used to estimate the energy exiting the ECal.

3.2.3.2 Hadronic Calorimeters

The HCal calorimeter is used to determine the energy of hadrons and consists of several systems distinguished by regions of η , namely the Tile Calorimeter, the LAr Hadronic end-cap Calorimeter (HEC), and the LAr Forward Calorimeter (FCal). This calorimeter is positioned around the ECal and is similar to the Ecal in that it has a barrel and two end-caps. It also has a FCal that is responsible for the flow of particles with low transverse momentum. The active medium in this component of the calorimeter is a scintillator and the absorbing material is steel. It covers a pseudorapidity range up to $|\eta| < 5.2$. Because the HCal is only reached by heavier particles than the ECal, the nuclear interaction length λ causes the particles to form larger and deeper showers. This measures the average distance the hadron travels before it undergoes the interaction.

Tile Calorimeter: Tile Calorimeter: Like the ECal, the Tile Calorimeter is used as a sampling calorimeter but is located outside the ECal and covers the range $\eta < 1.7$. However, the active and absorbing mediums are scintillating tiles and steel, respectively. This consists of a barrel and two extended barrel parts covering ranges of $|\eta| < 1.0$ and $0.8 < |\eta| < 1.7$, respectively [76]. This is primarily responsible for measuring the missing transverse momentum p_T^{miss} and provides information for the energy construction of the jets produced in pp collisions.

LAr Hadronic end-cap Calorimeter: The HEC has a pseudorapidity covering the range $1.5 < |\eta| < 3.2$. The active substance is liquid Argon, while the absorbing medium is copper.

LAr Forward Calorimeter: The FCal, unlike the HEC, has a range of $3.1 < |\eta| < 4.9$ and is 4.7 m from the interaction point in the z -direction. In addition to tungsten, FCal, similarly to HEC, uses liquid Argon as the active medium and copper as the absorber.

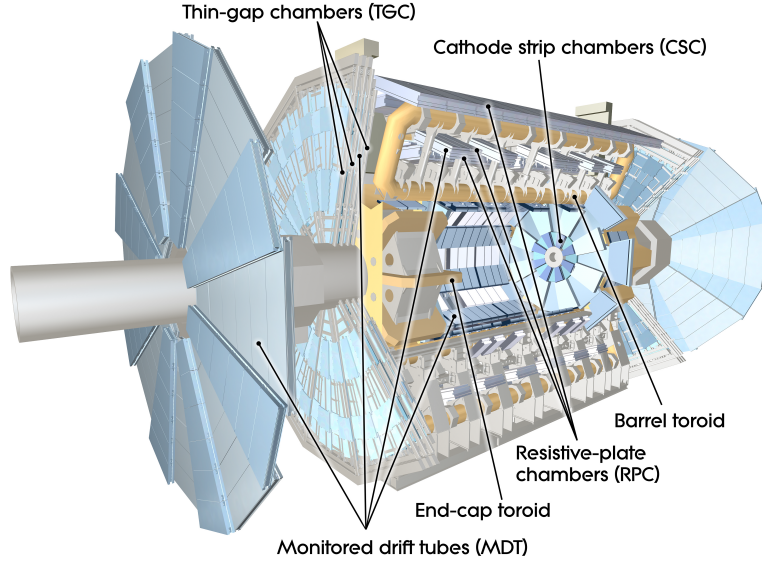


Figure 3.7: The ATLAS Muon Spectrometer [62].

3.2.4 Muon spectrometers

The muon spectrometer (MS), seen in Figure 3.7, is the outermost layer of the ATLAS detector. This part of the detector tracks muons because they are the only charged particles that cannot be stopped by the calorimeters due to their large mass and lack of strong interactions, so they pass through the calorimeter layers with little energy loss. Other particles that are fully contained before they reach the outer radius of the hadronic calorimeter are an exception. The MS consists of high-precision tracking chambers, trigger chambers, and superconducting air-core toroid and magnets [77]. It primarily identifies and reconstructs muons with a pseudorapidity in the range of $|\eta| < 2.7$ and p_T range 3 GeV - 1 TeV, as well as triggers events with muons in the range of $|\eta| < 2.4$. Like the ECal, the MS consists of two parts: a barrel and two end caps. The barrel covers a range of $|\eta| < 2.4$ for pseudorapidity, while the latter covers a range of $1.05 < |\eta| < 2.7$ for pseudorapidity. Therefore, it helps to distinguish muons from other charged particles. Figure 3.7 also shows the four detectors of the MS system, which include Monitored Drift Tubes (MDT), Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC), and Thin Gap Chambers (TGC). These modules play a key role in the detection process and use a variety of detection techniques. RPC and TGC cover different pseudorapidity ranges, with RPC $|\eta| < 1.05$ and TGC $1.0 < |\eta| < 2.4$. While CSC covers $|\eta| < 2.0$, MDT covers the entire pseudorapidity range of the spectrometer. TGC and RPC are responsible for triggering events and measuring the position of the muon track. MDT and CSC are responsible for the precise measurement of the path of the muons through the MS. The muon momentum is measured by the magnetic field generated by the three large superconducting toroidal magnets with an air core.

Monitored Drift Tubes MDT covers the entire pseudorapidity range of the spectrometer. MDT is responsible for the precise measurement and identification of muon

tracks in the barrel part of the detector. It consists of long cylindrical tubes filled with a gas mixture, typically a noble gas such as argon or a mixture of argon and carbon dioxide. Each tube is equipped with a central anode wire running along its entire length and is surrounded by a series of cathode strips. When a charged particle passes through the gas volume, it ionizes the gas atoms creating positive ions and free electrons.

Cathode Strip Chambers CSCs are made up of arrays of crossed negatively and positively charged copper cathode strips and anode wires in a gas volume. The anode wires experience an avalanche of electrons when muons travel through them and knock electrons off the gas atoms, which then flock to the anode wires. At an angle to the wire direction, positive ions travel away from the wire and toward the copper cathode, creating a charge pulse in the strips. CSCs are responsible for providing precise tracking, momentum measurement, and spatial resolution for charged particles.

Resistive Plate Chambers RPCs are made up of two parallel plates with a positively charged anode and a negatively charged cathode. Both plates are made of plastic with extremely high resistivity, and a thin volume of gas separates them. The electrons in the gas atoms are knocked out when a muon passes through the chamber. These electrons then collide with other atoms, creating an electron avalanche. RPCs are responsible for precise timing measurements, efficient triggering and high detection performance.

Thin Gap Chambers TGCs consist of closely spaced parallel electrode plates with a small gap between them. The gap is usually filled with a gas mixture, such as a combination of argon and carbon dioxide. The electrode plates are made of a conductive material, such as metal, and are separated by thin insulating layers. TGCs are known for their fast response time, typically on the order of nanoseconds, and their ability to handle high particle rates. The small gap between electrode plates allows for efficient signal amplification and fast charge collection, enabling TGCs to accurately measure the timing of particle interactions.

3.3 ATLAS Trigger System

The ATLAS trigger system is responsible for deciding whether to keep or discard an event. During the first long shutdown after Run 1, this system was extensively upgraded to cope with the high collision rates and high luminosity. As a result, it is implemented in both software and hardware, so it can be used both online and offline. The online data processing is done while the data is being recorded, while the offline data processing is done after the event is recorded. The ATLAS trigger system and data acquisition are shown schematically in Figure 3.8. The collisions occur at a revolution frequency driven by the luminosity as described before. This produces a significant amount of events from which events physicists are not interested in are discarded [79]. The hardware component consists of a Level 1 (L1) trigger. This trigger is responsible for analyzing event data at a rate of 40 MHz while accepting events at 100 kHz for further processing. It also identifies objects with high E_T by using information from MS and calorimeters. The software component includes a high-level trigger (HLT). The HLT uses the information

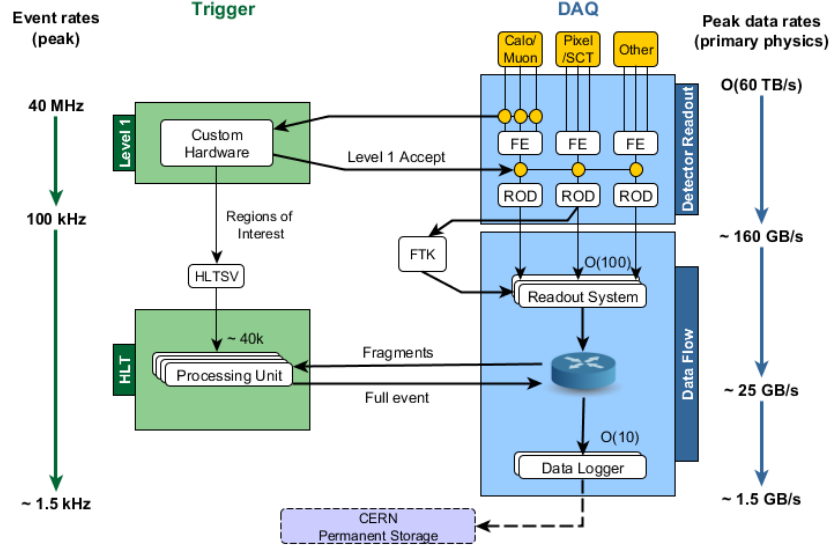


Figure 3.8: Schematic representation of the ATLAS trigger and data acquisition systems after Run 2 [78].

it receives from L1 and selects events at a rate of 1 kHz. As this information is gradually reduced, it is further written to data memory and then sent for full reconstruction. All of these levels are interconnected. The next level refines the decision of the previous level and applies additional selection criteria as needed. In addition to the two levels, there is also an event filter called trigger chains. This consists of a list of triggers that come into play during the event selection process. The number of events is reduced from the typical 40 MHz to 100 KHz by the L1T and 1.5 kHz by the HLT, reducing the data from 60 TB to 160 GB and 1.5 GB.

3.4 Event Reconstruction in ATLAS

Raw detector signals are converted into measurements of the particles produced by the collisions during event reconstruction. This is done in several steps using reconstruction algorithms before the collision events can be used for further analysis.

The particles produced in the pp collisions deposit energy in the detector after leaving the collision point. This energy deposit is then measured and used to define physics objects such as photons, electrons, muons, jets, and missing transverse momentum. This is accomplished by using reconstruction algorithms that reconstruct the physics objects and measure the outgoing particles produced during the collision. The reconstruction algorithms from the low-level detector signals such as MS tracks or energy depositions in ECal or HCal. These algorithms play a very important role in the accurate identification of physics objects after the reconstruction step to ensure that the objects are traced

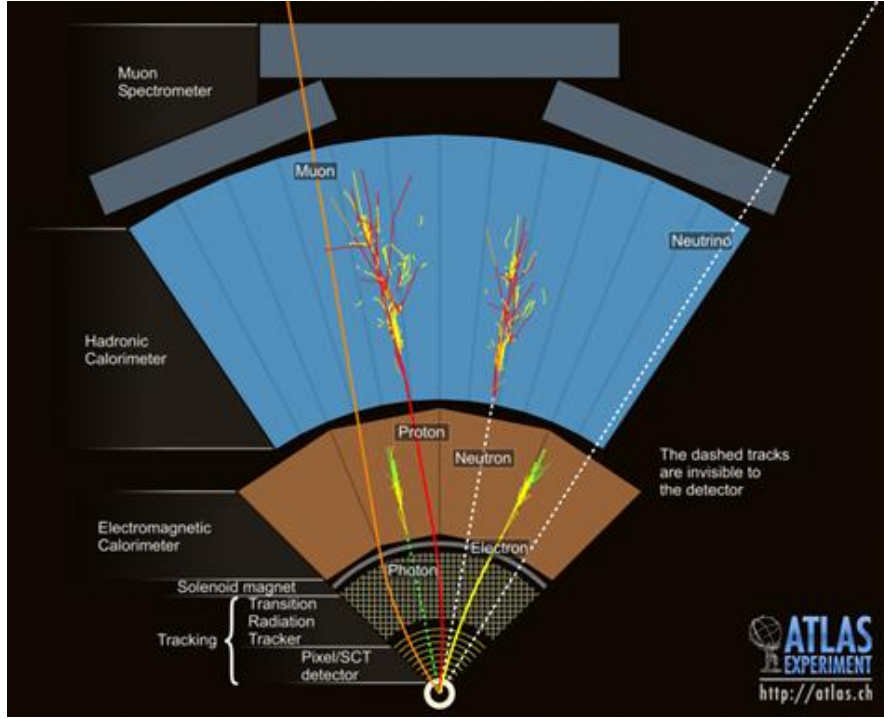


Figure 3.9: Schematic representation of different particles interaction with the ATLAS subsystem [80].

throughout the event. This step is very important because the reconstructed photons could be misidentified from electrons, quarks, or gluons instead of from photons. Therefore, the identification and isolation steps are very important because they ensure that the efficiency of the background is low and the efficiency of the reconstructed objects is high.

3.4.1 Track Reconstruction

Charged particles passing through the ID leave hits in the subdetectors that yield tracks. These tracks yield the trajectories of the charged particles [81] and are reconstructed by implementing the inside-out and outside-in algorithms [82]. Primary pattern recognition starts in the Pixel detector and works its way to the outer edge of ID in the inside-out algorithm, as the name implies. The first step is seeding, which occurs in the silicon tracker, and the second step is hit discovery, which occurs in the silicon tracker. The outside-in method, on the other hand, is responsible for improving particle tracking when they decay at a displaced vertex and works in the opposite direction as the inside-out approach. The algorithm starts with hits on the TRT and works its way back to the Pixel detector.

Charged particle tracks, also known as trajectories, are critical to the reconstruction of many physics objects. Preprocessing, track-finding, and post-processing are the three stages of the track reconstruction process. This process reconstructs charged tracks us-

ing data from the Inner Detector [83, 84]. In the preprocessing phase, raw data from the Pixel and SCT detectors are received and converted into clusters. While the information from TRT and SCT is received, the timing information and clusters are then converted into drift circles and space points, respectively.

In addition to the azimuthal angle and polar, the tracks can also be defined by the minimum distance in the longitudinal direction and the minimum distance in the transverse plane from the center of mass. These parameters can be represented by z_0 and d_0 , respectively, giving the definition of tracks the following representation:

$$(d_0, z_0, \phi\theta, \frac{q}{p}), \quad (3.9)$$

where $\frac{q}{p}$ is the ratio of charge to momentum.

3.4.2 Electron Reconstruction

As mentioned in Section 3.2.3, the ECal is responsible for measuring the energy of electrons and photons [85]. This is where electrons deposit their energy. Figure 3.9 shows the interactions of each subdetector with different particles. These particles leave showers in the respective calorimeter. The electrons and photons leave showers in the ECal, while hadronic particles leave showers in the HCal and muons pass through all these calorimeters and register the showers in the MS.

Electron reconstruction is performed by matching energy deposits in the ECal with traces formed in the ID. This process consists of four distinct steps, namely seed cluster reconstruction, trace reconstruction, electron-specific trace matching, and electron candidate reconstruction [85]. The seed-cluster reconstruction implements a sliding-window algorithm [86] with a size of 3×5 towers to search for energy clusters with $E_T > 2.5$ GeV. Some of the reconstructed electrons are jets faking an electron, and the likelihood approach is used as an identification technique [85]. The identification is classified into loose, medium, and tight.

3.4.3 Jet Reconstruction

The most commonly used algorithm for reconstructing jets in ATLAS is the anti- k_T clustering algorithm [87] with the maximum allowed distance parameter of $R = 0.4$. This reconstruction takes place in the EM and hadronic calorimeters. These jets are defined by energy deposits left by hadrons in ECal and HCal. Moreover, they must be within $|\eta| < 4.5$. However, the p_T requirement depends on the position of the jet. Thus, the central jet must be within $|\eta| < 2.4$ and the outer region must have $p_T > 25$ GeV and $p_T > 30$ GeV, respectively.

3.4.4 Muon Reconstruction

The muon reconstruction uses information from ECal, HCal, and MS for the reconstruction. However, it is the information from MS that is used to create a muon object [88]. The muon objects are reconstructed from different types of muons that depend

on different information from the subdetectors, depending on the requirements of the track. These types are the combined muon, the segment-tagged muon, the calorimeter-tagger muon, and the extrapolated muon [88]. The combined muon is reconstructed by combining the tracks from MS and ID and re-matching them with all hits, while the segmented-tagged muon is reconstructed by extrapolating an MS segment to the ID, finding a suitable track candidate. In contrast to the segmented-tagged muon, the calorimeter-tagged muon is reconstructed by extrapolating an MS segment to a calorimeter cluster. This is used only in regions where MS is only partially available. Finally, the extrapolated muons are those that have only tracks in MS but do not match any track in ID.

3.4.5 Photon Reconstruction

The calorimeter is divided into a grid of towers, each of which measures $\Delta\eta \times \Delta\phi = 0.075 \times 0.025$, and the combined tower energy is equal to the sum of the energy deposits from all layers of the calorimeter. Photon reconstruction is initiated using a similar approach to electron reconstruction. Again, a sliding window algorithm is used during the cluster seed reconstruction step. This is done to find the seed, where clustering comes into play to reconstruct the clusters. A photon produced at an interaction point first passes through a tracker before entering an electromagnetic calorimeter, where it loses all of its energy through electromagnetic showers that can occur over a wide range within the calorimeter. Photon candidates must have $E_T > 2.5$ GeV to be reconstructed from energy clusters in the ECal [89]. Searching for energy clusters in the calorimeters is the first step in photon reconstruction. Cluster seeds with a total energy of $E_T > 2.5$ GeV or more are selected using a sliding window technique with a size of $\eta \times \phi = 3 \times 5$ towers. Multiple cluster seeds within any 2×2 tower range are selected using a second technique to eliminate duplicates, but only the one with the highest transverse energy is ultimately selected.

3.4.6 Missing Transverse Energy

There is a transverse momentum that is not detected in the detector and this is called missing transverse momentum, S_{ET}^{miss} . This is due to particles such as neutrinos escaping the detector without a measurable trace, making direct reconstruction impossible. This parameter measures the magnitude of the negative sum of the transverse momentum of the reconstructed physics objects and plays an important role in the physics analyses. It is determined using the final state particles from the hard pp interactions [90]. The S_{ET}^{miss} is given by the sum of the following terms [91]:

$$S_{ET}^{miss} = S_{ET}^{miss,e} + S_{ET}^{miss,\gamma} + S_{ET}^{miss,\tau} + S_{ET}^{miss,jets} + S_{ET}^{miss,\mu} + S_{ET}^{miss,soft}, \quad (3.10)$$

where,

- $S_{ET}^{miss,e}$ is the negative vectorial sum of the momenta of the electron
- $S_{ET}^{miss,\gamma}$ is the negative vectorial sum of the momenta of the photons
- $S_{ET}^{miss,\tau}$ is the negative vectorial sum of the momenta of the taus

- $S_{ET}^{miss,jets}$ is the negative vectorial sum of the momenta of the jets
- $S_{ET}^{miss,\mu}$ is the negative vectorial sum of the momenta of the muons
- $S_{ET}^{miss,soft}$ is the negative vectorial sum of the momenta of the soft term, which is detector signals that are associated with specific objects

Thus, the total S_{ET}^{miss} is finally given by:

$$S_{ET}^{miss} = \sqrt{(S_{Ex}^{miss})^2 + (S_{Ey}^{miss})^2}. \quad (3.11)$$

Chapter 4

Semi-supervised Learning-based search of heavy resonances decay to $\gamma\gamma$ with topological features

4.1 Introduction

In this chapter, we present the physics background and machine learning-based analysis for the search of resonances decaying to $\gamma\gamma$ with topological features at the LHC. For this purpose, a semi-supervised learning strategy in the form of weakly supervised learning is used. This approach is evaluated on the SM Higgs boson and background processes with identical decay products. The performance of semi-supervised learning is then compared to that of fully supervised learning.

As it comprised the last piece of the SM, the discovery of the Higgs boson at the LHC indicated that the SM had exhausted fundamental predictions. However, the SM has some shortcomings in explaining a variety of phenomena that have immense experimental evidence. This includes matter-antimatter symmetry. As a result, physicists are working ceaselessly to conduct experiments dedicated to exploring physics beyond SM.

Due to the large amount and complexity of data produced at the LHC, advanced anomaly detection methods are required. This is where the use of ML techniques comes into play. Such methods have been proven useful in the discovery of the Higgs boson. The search for new resonances is critical to the overall goal of uncovering new physics beyond the SM. Therefore, the absence of signals in inclusive searches means that new resonances, if hidden in the data, will be modest.

ML has the potential to help solve this challenge. This is evident in fully supervised learning, which has proven effective in exploring corners of phase-space. The phase spaces where the SM Higgs boson was discovered are an excellent example. The problem with this strategy is that it is hampered primarily when the new physics populates the phase-space in a different way than the baseline model. The use of semi-supervised learning techniques, especially weakly supervised learning, based on sidebands that are

treated as the absence of a signal provides an advantage in detecting topological differences between the prospective signal region and the sidebands. Due to the complexity of the phase-space under study and the way it is populated with backgrounds and possible signals, this approach may appear inefficient. The SM Higgs boson is used in this study to quantify this effect.

The ATLAS and CMS experiments at the LHC have led to the identification of a number of anomalies based on differences in lepton production [6–16]. The number of final states, such as dileptons with opposite and same sign, three leptons in the presence and absence of b-quarks, are among them. It is worth noting that these final states arise in regions of phase space where various SM processes are dominant.

The proposed hypothesis, composed of two new scalar fields, H and S , can explain these anomalies [6, 10, 16]. These scalar fields have masses around 270 GeV and 150 GeV, respectively. This provides an incentive to search for Higgs-like resonances with masses around the electroweak (EW) scale, where SM backgrounds are large. Topological criteria are key in suppressing backgrounds in this case.

4.2 Application of Machine Learning in HEP

ML is a branch of artificial intelligence (AI) that is becoming more prevalent in a variety of computer systems. Algorithms in this field are able to extract insights from data without being explicitly coded. Because the programmer has no control over what happens within the algorithm during the learning phase, ML is referred to as a “black box”. These algorithms use the learned insights to generate a new set of rules, which are then applied to new data that the algorithm has never seen before to make generalizations. The nature and understanding of the problem to be solved, along with the structure and format of the data, are some of the factors that determine which algorithm to apply. In addition, most of these algorithms have been proven successful in HEP regression and classification problems [92].

The use of these algorithms in HEP began in the 1990s [93], with one of their most notable contributions being the discovery of the Higgs boson at the LHC in 2012 [18, 19, 94, 95]. Their participation usually takes place during the intermediate steps of data processing [96, 97], with observations being performed during the final steps of the measurement. Therefore, they are mainly used for the classification of signal and background events in HEP search [98]. These algorithms have been shown to be beneficial as they improve the search significance when compared to the standard rectangular cuts search approach [99]. However, the latter implies a cost-benefit tradeoff. The work presented here is the search for resonance with topological features at the LHC, which is compared to Higgs boson production. The semi-supervised learning technique in conjunction with DNNs is used to achieve the goal of this study.

For particle search, algorithms such as Boosted Decision Trees (BDTs) [89, 100–107] and artificial neural networks (ANNs) [108–110] were widely used. According to Ref. [111],

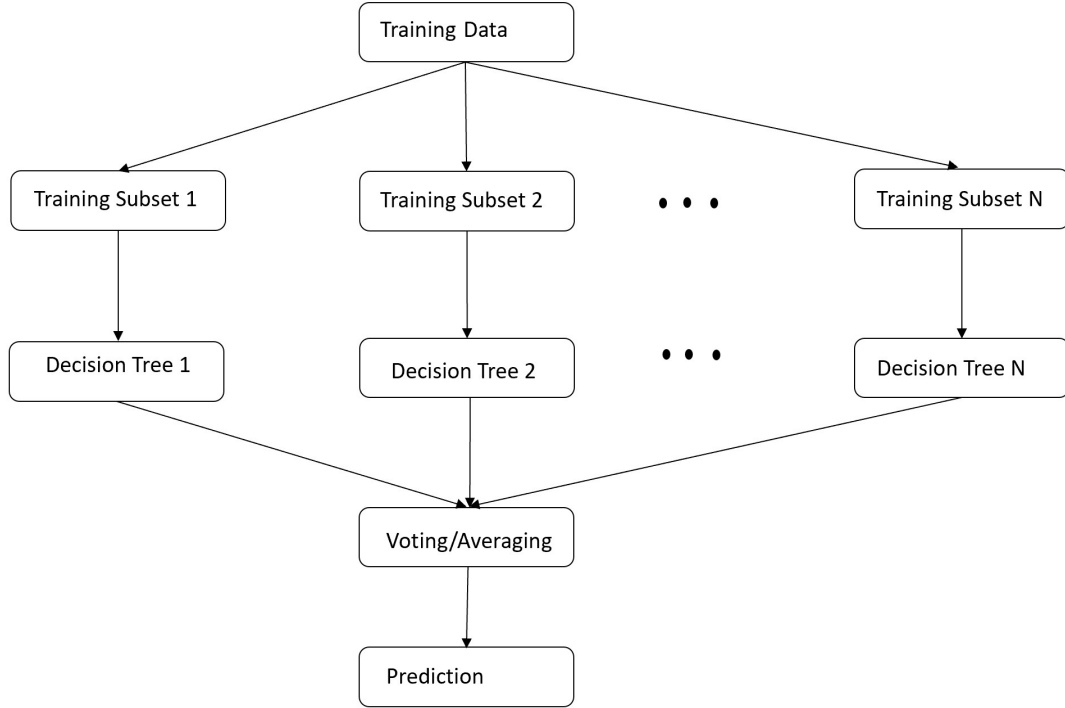


Figure 4.1: A schematic illustration of a Random Forest comprised of N number of decision trees.

ANNs can approximate the correlations between input variables. In evaluating the two approaches, ANNs outperformed BDTs when only four input variables were used [110]. According to Ref. [112], on the other hand, BDTs outperformed ANNs when the number of input variables was increased. The use of BDTs was extended to software triggers [113, 114], while the use of ANNs was extended to electronics [84], bottom quark tagging [115], and tau identification [116]. Since they consist of a set of trained classifiers, BDTs outperform decision trees (DTs) in terms of performance. Ref. [117] investigated the use of generalized models and Gaussian mixture models, in addition to BDTs and multilayer perceptrons (MLPs). The performance of both generalized models and Gaussian mixture models was similar to that of BDTs and MLPs.

4.2.1 Random Forest

Random Forest (RF) is one of the most important algorithms in ML, introduced by Leon Breiman in 2001 [118]. This algorithm is widely used for classification and regression tasks in industries such as healthcare [119, 120], banking [121, 122], and e-commerce [123, 124]. The basic components of RF are DTs built on subsets of the training data and then combined into an ensemble that performs majority voting to make a final decision. Figure 4.1 shows an illustration of RF, which consists of N DTs. These trees make predictions for the classification task, with the class with the most votes forming the prediction of the model. The intuition behind this is that a large

number of models that are not even correlated with each other work as a committee, and this is somehow expected to outperform individual DT models.

Unlike DT, which is prone to overfitting, RF can avoid this problem because it is built from uncorrelated DTs. In addition, RF differs from BDT in the sense that the results of the DTs are aggregated at the end of the training process, whereas in BDT the process occurs during training, with each DT being part of the decision. The decorrelated trees in RF are trained independently on different subsets of the training data [125]. The method uses a bootstrap aggregation approach, one of the purposes of which is to reduce variance using the resampling with the replacement method [126]. This technique uses different methods for decision-making, i.e., averaging the decision trees in the case of a regression problem and applying a majority rule in the case of a classification task.

Being composed of a set of DTs, RF is known to achieve better prediction accuracy than a single DT [118]. In the initial phase of RF, training data consisting of labels and features are input into different DTs. These features are randomly selected when the nodes are partitioned. RF is based on ensemble learning, which relies on combining multiple DT classifiers to achieve better performance. These trees are trained on different subsets of the training data, taking the average or using the majority rule to improve performance rather than relying on a single DT. The nodes of each DT in the RF serve as attributes, while the leaves serve as decision-makers of the classifier. The nodes are created by selecting the best feature from the subset of features used to train that specific DT.

Moreover, the quality of the decision for each node is determined by calculating the entropy gain or Gini index for each feature. The RF can be optimized by controlling parameters such as the number of DTs that make up the RF, the maximum depth of each DT, and the maximum number of features used to train each tree. The DTs are evaluated independently and finally come together to make the final decision by majority vote.

The RFs have found their application in HEP, this includes but is not limited to particle detection [127], data quality investigation [128], in addition to their intensive implementation in both the D0 Collaboration [129–132] and the BaBar Collaboration [133, 134]. The DNNs managed to perform better than RF in Ref. [135]. The advantages of RF are also that they are less affected by noise in the data and can handle missing values. Each structure of DT consists of leaves and nodes that represent the features in the subset of training data and the decisions, respectively. Since it is known that DTs tend to overfit, the problem of overfitting does not exist in the case of RF. Moreover, RF provides more accurate predictions compared to DTs, which is of course due to its structure where a committee of DTs works together towards the same goal.

The use of ML, in this case, was motivated by fully supervised learning, in which the production mechanisms, topologies, and signatures were fully described by theory, and the mass of the SM Higgs boson was the sole free parameter in the search. The search for new resonances, including new Higgs-like bosons, on the other hand, may entail subtle topologies. Fully supervised learning becomes onerous and ineffective in this

situation. Such research may be carried out with the use of a Lagrangian based on effective interactions, which requires a large number of parameters. As we move away from fully supervised learning and toward full-blown weakly supervised learning, this study evaluates different approaches.

4.2.2 Deep Learning

Deep learning is part of the ML category. Image recognition [136–138], object detection [139], face recognition [140], speech recognition [141], and machine translation [142] have all benefited from this paradigm. DNNs are a type of DL algorithm. The architecture of a DNN consists of numerous layers stacked together. These layers are responsible for converting the outputs of the previous $(k-1)^{th}$ layer into inputs for the next (k^{th}) layer. Equation 4.1 represents the mathematical formulation of this paradigm:

$$h^{(k)} = f^{(k)}(b^{(k)} + W^{(k)}h^{(k-1)}), \quad \vec{h}^{(0)} = \vec{x}, \quad (4.1)$$

where $W^{(k)}$ denotes the weights matrix, $b^{(k)}$ denotes the bias term, $h^{(k-1)}$ is the output of the $(k-1)^{th}$ hidden layer, $f^{(k)}(\cdot)$ denotes the activation function of the k^{th} layer, and $\vec{x}$ denotes the input vector [143].

The depth of a DNN is determined by the number of hidden layers in the network, which can be classified as shallow or deep neural networks. Shallow neural networks have a single hidden layer, while deep neural networks have more than two. On the other hand, shallow neural networks have some limitations when it comes to discovering effective features in high-dimensional data [144]. As a result, they have difficulty learning complex nonlinear functions [145]. The evolution of ANNs into DNNs, as described in Ref. [93], has led to an increase in the training and processing capacity of the algorithms.

Unlike shallow neural networks, DNNs have a structural design that makes them more effective and able to learn complex models [146, 147] as network depth increases [148]. According to Ref. [149], their success can be seen in HEP, where DNNs outperformed shallow ANNs. Consequently, they are useful for learning discriminative information in physics variables [145].

The complexity and huge amount of data generated by LHC experiments have led to a significant increase in the use of ML techniques in this field. DNNs have improved classification performance by 8% compared to previous approaches, according to studies such as Ref. [145]. They also proved useful for searching for exotic particles in collider, and their performance was comparable to or better than typical jet variables [150]. However, they have a vanishing gradient problem that can be solved using some of the strategies described in Section 5.4.3. Other DL approaches, such as convolutional neural networks (CNNs), have been used to distinguish quark and gluon jets in Ref. [150], and they appeared to perform better than typical observables designed by physicists. It is also noted that DNNs can extract reliable physics information from poor simulations and that they are also used for image identification such as jet substructure detection [151].

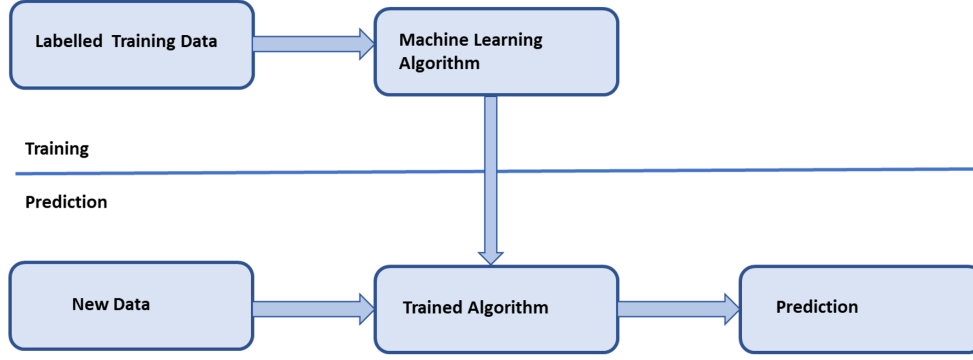


Figure 4.2: Schematic representation of supervised learning.

4.2.3 Fully Supervised Learning

ML algorithms take a set of input vectors $\vec{x}_i$ of dimensionality n . These vectors are generally referred to as features. The features are passed to an algorithm with the goal of learning a function that allows the algorithm to generalize the learned information to previously unseen data. Various parameters of the algorithm are set to facilitate this learning process. When methods such as ANNs [152, 153] and Support Vector Machines (SVMs) [154] are used instead of heuristic selectors [155], fully supervised learning has been shown to be successful in event selection in HEP. Given a binary classification task, the basic goal of fully supervised learning is to train fully labeled data with the goal of finding a mapping function $f(x)$ that maps the input $\mathbb{R}^d$ to a set of classes $\mathbb{K} = \{1, \dots, k\}$, where each example $\vec{x}_i$ has a label $y_i \in \{0, 1\}$.

Figure 4.2 shows a schematic representation of all the steps involved in fully supervised learning. The algorithm is trained with fully labelled data and used to predict new data. The goal of the training is to find a mapping from x to y while minimizing the loss function, which can be expressed as binary cross-entropy:

$$\ell(y, \hat{y}) = -y \cdot \log \hat{y} + (1 - y) \cdot \log(1 - \hat{y}), \quad (4.2)$$

where $\hat{y}$ is the model output and y denotes the desired outcome. The main goal is for a trained model to find a function to estimate labels in an unknown dataset. Minimizing the loss function helps in finding the best solution [156]:

$$f_{full} = \underset{f: \mathbb{R}^n \rightarrow [0,1]}{\operatorname{argmin}} \sum_{i=1}^N \ell(y_i, \hat{y}_i), \quad (4.3)$$

where f is the prediction function, $\hat{y}_i$ is the i^{th} model output, y_i is the i^{th} target output and ℓ is the loss for a single example [157]. During the process, the model generates its own output while adjusting the weights using a gradient descent algorithm in the form of backpropagation. This is repeated until the output of the model approaches the expected output. The fully supervised learning paradigm is referred to as “*learning with a teacher*” [158] because of this stated approach. Fully supervised learning is mostly used for regression and classification problems, and training algorithms for this learning paradigm require a large amount of labeled data, especially when implemented on DNNs.

4.2.4 Semi-supervised Learning

Unlike fully supervised learning, unsupervised learning does not require labeled data and expects the model to learn from itself based on similarities in the input data [159]. Semi-supervised learning [160, 161] is the result of a combination of supervised and unsupervised learning. Therefore, both labeled and unlabelled data are used in this learning paradigm. Therefore, both labeled and unlabelled data are used in this learning paradigm. In semi-supervised learning, the model is first trained on the small amount of accessible labeled data. Then, the partially trained model is trained to generate labels for the unlabelled data. The model is then trained on a mixture of labeled and pseudo-labeled data. During training, this mixture ensures that error is reduced and accuracy is improved.

On a classification problem for breast cancer prediction, Ref. [162] compared the performance of fully supervised and semi-supervised learning. This was tested on a number of algorithms, including logistic regression, k-nearest neighbor (KNN), and DT, and the results showed that the semi-supervised learning strategy produced results quite similar to those of fully supervised learning. Speech analysis, web content classification, and text document classification are examples of semi-supervised learning applications. Due to a large amount of data that requires labeling, such as audio data, website material, and text data, this is an appropriate technique for the above areas. All of these tasks require a significant amount of human intervention.

4.2.4.1 Weakly Supervised Learning

DL models require a large amount of data and a lot of computational power. The labeling of such training data by human specialists can also be expensive. If done manually, this can be potentially unreliable [163]. Because of the difficulty of collecting large amounts of labeled data, many researchers have explored the paradigm of weakly supervised learning, particularly in robotics, disaster resilience, medical diagnostics, bioinformatics [164], and computer vision for action identification [165] and object recognition [166]. As a result, interest has grown in HEP [1, 156] for quark vs gluon tagging [156].

Weakly supervised learning is considered a cheap form of supervision [163, 167] because it is designed to learn from data with imprecise labels [168] that are treated as ground truth. Compared to fully supervised models, the performance of this strategy on visual detection was better in Ref. [169]. Sentiment analysis [170] and retrieval of complex information from trained models [171, 172] are two other applications. The weakly supervised learning comes in three self-explanatory forms, namely [173]:

- Incomplete weak supervision: when labels are assigned to only a subset of the training data,
- Inaccurate weak supervision: a situation where certain examples are assigned incorrect labels, and
- Inexact and ineffective supervision: in this scenario, only coarse-grained labels are assigned.

One of the main advantages of weakly supervised learning is that the time required to manually label the training data is minimal. Ref. [174] has succeeded in saving 6.8% of the time spent on manual labeling of training data for supervised learning methods. Moreover, weakly supervised approaches were equal to fully supervised learning in terms of performance [92, 156, 166, 174–176]. The following equation is used to optimize this weakly supervised learning:

$$f_{weak} = \underset{f: \mathbb{R}^n \rightarrow [0,1]}{\operatorname{argmin}} \sum_{K's} \ell \left(\frac{1}{|K|} \sum_{i \in K} \hat{y}_i, y_K \right), \quad (4.4)$$

where K 's denotes the training data batches and y_K is the signal fraction in each batch [157].

Weakly supervised learning can be considered as a subset of semi-supervised learning [177]. This subfield was developed to enable models to learn from partially labeled data, and it derives from the lack of fully labeled data. Studies such as Ref. [168] use outputs from unsupervised models to generate labels for weakly supervised learning models. Ref. [178] combines weakly supervised learning and fully supervised learning by training two different ANNs based on these approaches. The two models are called the target network and a confidence network. The former is trained on weakly labeled large data, while the latter is trained on small and fully labeled data. The confidence network is responsible for updating the weights of the target network so that the weights no longer need to be computed directly from the weak labels. The equation 4.2.4.1 is used to compute this:

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \frac{\eta_t}{b} \sum_{i=1}^b c_{\theta}(x_i, \tilde{y}_i) \nabla \mathcal{L}(f_{\mathbf{w}_t}(x_i, \tilde{y}_i)), \quad (4.5)$$

where η_t is the global learning rate, $\mathcal{L}(\cdot)$ is the loss function, $c_{\theta}(\cdot)$ is the learned evaluation function of the confidence network, and $\tilde{y}_i$ is the noisy label, while t and b denote a time step and the batch size, respectively. The weakly supervised learning strategy is of interest to academics and they use this space to increase the performance of models at a low cost.

4.2.4.2 Supervised learning with unlabelled data

Several studies [179–184] have been conducted to improve the performance of fully supervised learning algorithms with unlabelled data. In this example, the training data primarily consists of large u unlabelled samples $\{X'_1, \dots, X'_u\}$ with limited l labelled samples $\{(X_1, \theta_1), \dots, (X_l, \theta_l)\}$. To learn from the training data containing both labeled and unlabelled samples, a combination of statistical approaches such as expectation maximization (EM) and a naive Bayes classifier can be used [181]. This is done to train the classifier on labeled data while using it to assign labels to unlabelled data. To achieve this goal, DL techniques have also been used [180], and this approach is called supervised learning using labeled and unlabelled data (SULU) [183].

Ref. [183] has proven that using the SULU algorithm instead of the EM algorithm leads to better results. On the other hand, the performance of the EM algorithm improves as the number of labeled samples increases. SULU, on the other hand, requires only a small amount of labeled data to perform better and has consistently outperformed algorithms trained with only labeled data. Ref. [179] co-trained two separate fully supervised learning algorithms on labeled data. Statistical approaches are also used to sample the unlabelled data, with one of the trained algorithms assigning labels to the other algorithm. The process is performed repeatedly to label the unlabelled data, while high classification confidence examples are added to label the data, which is then used to re-train the model [185].

The co-trained algorithms create weak predictors that are then bootstrapped using the unlabelled data, as described in Ref. [186]. In contrast, we introduce “guided supervised learning”, a technique that mixes multiple signal production mechanisms and assigns them the same class label. The method takes into account the fact that the combined mechanisms are fundamentally different. The goal of this research was to find out whether an ML algorithm can distinguish each production technique during the testing phase.

4.3 Dataset and event selection

The simulated Higgs to di-photon, $pp \rightarrow h \rightarrow \gamma\gamma$ events are used to test the ML approaches used in this study. This research focuses on proton-proton collisions at the LHC with a center-of-mass energy of 13 TeV. The background and signal samples were created using the `MadGraph5_aMC@NLO 2.6.7` framework. This framework is designed to give all of the elements needed for the phenomenology of the SM and BSM. Calculations of cross-sections of effects, the formation of hard events, and their matching with event generators are some of these elements [166].

It is also worth noting that the events in this study were generated in QCD with next-to-leading order (NLO) accuracy. `PYTHIA 8.2` was used to simulate parton showering and hadronization in conjunction with `MadGraph5_aMC@NLO 2.6.7`. The ATLAS A14 event tuning and NNPDF2.3 LO parton distribution function set were used for this purpose [187]. `Delphes 3`, a modular framework for fast simulation of the ATLAS experiment, was used for event processing [188]. Finally, by applying a set of generator-level cuts, events of interest were simulated using the following selections:

- The transverse momentum of the photons must be greater than 25 GeV,
- The di-photon invariant mass must be between 105 and 160 GeV.
- Hadronic jets were reconstructed using the anti- k_t technique [189] implemented in `FastJet 3.2.2` [190] package was implemented, with radius parameter $R = 0.4$.
- Jets originating from bottom quarks are identified as b -jets using b -tagging algorithms [191],
- Jets with $p_T > 30$ GeV and $|\eta| < 4.7$ are considered.

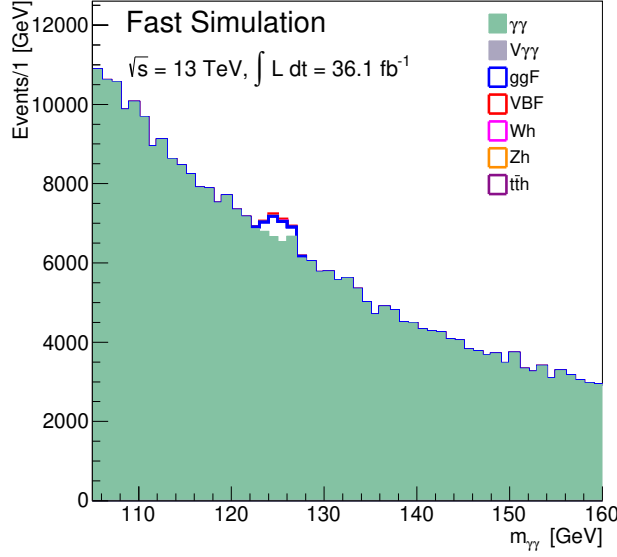


Figure 4.3: The spectrum of the invariant di-photon mass at 13 TeV of the proton-proton collision corresponds to an integrated luminosity of 36.1 fb^{-1} . The cross sections times the branching ratio of the di-photon decays are used to standardize the different production modes of the Higgs boson samples of the SM

Process	Entries	Event Yields
$\gamma\gamma$	96447	323436
$V\gamma\gamma$	26613	467
ggF	224835	1605
VBF	240085	129
Wh	281209	41
Zh	169312	23
$t\bar{t}h$	160265	18

Table 4.1: The expected events are normalized to the number of entries for each process.

- Jets in a cone of size $R = 0.4$ are deleted.
- Electrons and muons must have $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$.
- Finally, at least two photons must have $p_T > 25 \text{ GeV}$ and $|\eta| < 2.37$ selected from an inclusive event selection.

Non-resonant $\gamma\gamma$ and $V\gamma\gamma$ ($V=W, Z$) processes are the two primary backgrounds considered for this study. More than 80% of the total background contribution comes from the non-resonant. Different Higgs boson production methods, such as gluon gluon fusion (ggF), vector boson fusion (VBF), and associated productions with a vector boson (Vh) or a top-antitop quark pair ($t\bar{t}h$), have been considered in the signal samples. The

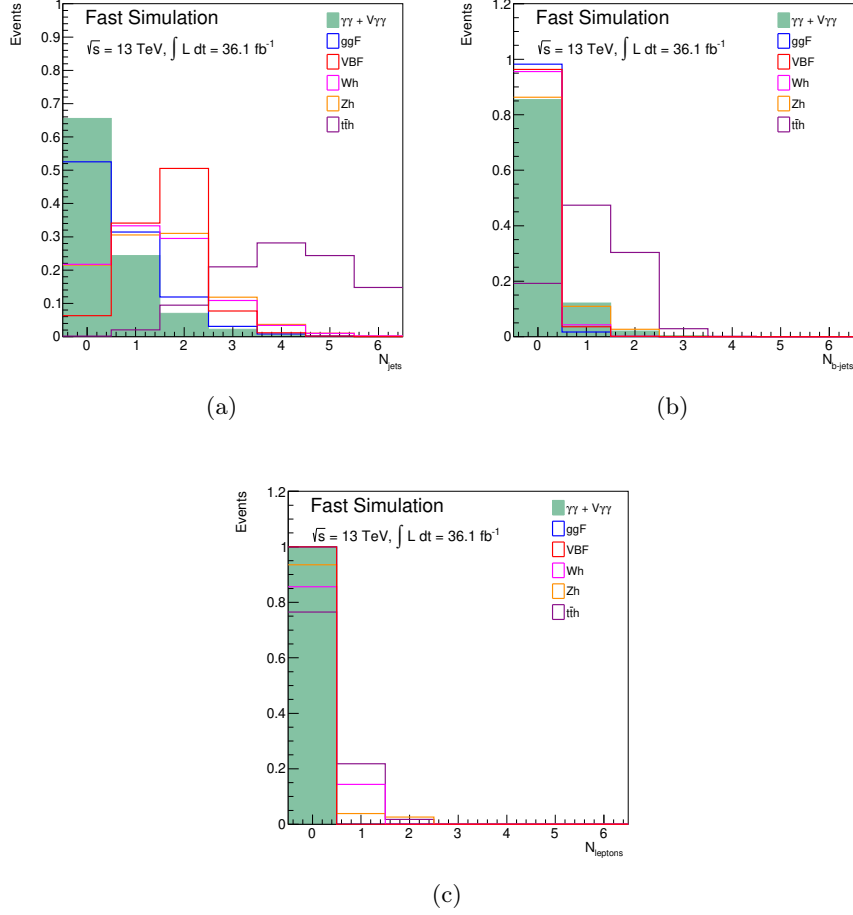


Figure 4.4: Signal and background distributions of the object multiplicity: (a) number of reconstructed jets per event, (b) number of b -jets per event identified by the b -tagging algorithm, and (c) number of reconstructed leptons per event

samples from these different processes are then normalized using the recommendations of the Higgs cross-section working group [192]. As shown in Table 4.1, the number of entries for each process is normalized to the predicted event yields. For both training and testing purposes, the data are then split into 50%:50% samples. Ref. [193] provides the $h \rightarrow \gamma\gamma$ branching ratio ($\approx 0.23\%$). The plot of the reconstructed di-photon invariant mass spectra, $m_{\gamma\gamma}$, for both signal and background, is shown in Figure 4.3.

4.3.1 Kinematic features

This research focuses on a total of 13 distinguishing features of particle physics describing ATLAS detector objects. In addition to using the phase-space associated with each event, these features are used to identify the background and the various SM Higgs boson generation mechanisms. These features also include information about object multiplicity. As shown in Figure 4.4, this includes the number of reconstructed jets,

b -jets, and reconstructed leptons. These kinematic features were also used to help the classifier distinguish between hadronic and leptonic decays and to isolate mechanisms that lead to the presence of at least one b -jet in the final states, such as $t\bar{t}h$.

To isolate processes with high transverse energy, such as Wh and $t\bar{t}h$, the following kinematic properties (shown in Figure 4.6) are used:

- The reconstructed missing transverse energy,
- The azimuthal angle separation between the leading jets and the missing transverse energy,
- The azimuthal angle separation between the di-photon system and the missing transverse energy, and
- The azimuthal angle separation between the di-photon system and the leading jets.

The jet pseudorapidity, η_{jet} , is used to distinguish processes that have at least two jets in the central region ($t\bar{t}h$) or forward region (VBF, Wh and Zh) of the detector. The transverse momentum p_{jet}^T is used to distinguish between processes with hard or soft jets. In addition, features such as di-jet mass, m_{jj} , and the pseudorapidity separation, $\Delta\eta_{jj}$, are used to distinguish VBF from other processes. Figure 4.5 shows the distributions of these features.

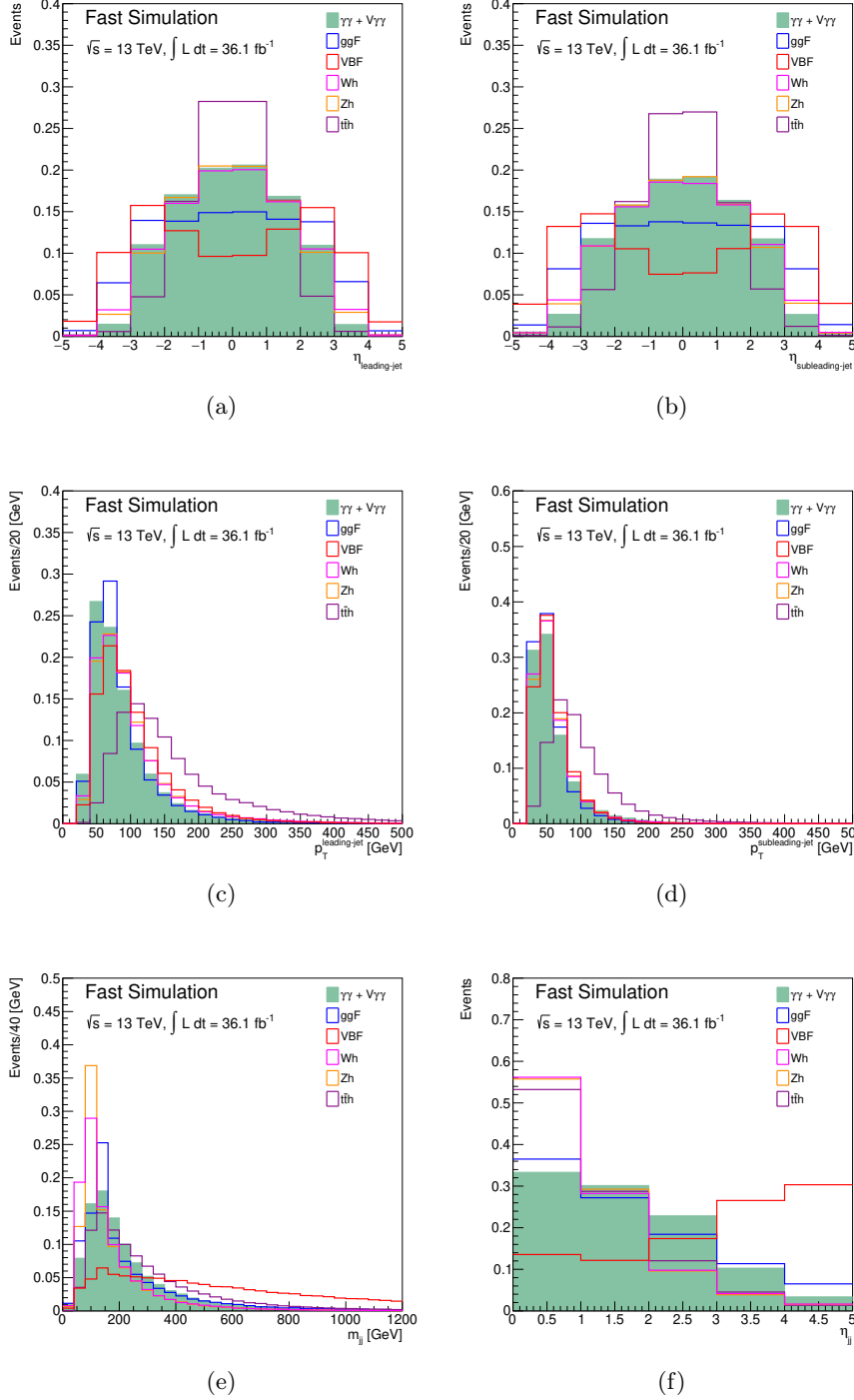


Figure 4.5: Signal and background distributions of (a) pseudorapidity of the leading jet, (b) pseudorapidity of the sub-leading jet, (c) transverse momentum of the leading jet, (d) transverse momentum of the sub-leading jet, (e) pseudorapidity separation of the leading and sub-leading jet, and (f) invariant mass of the leading and sub-leading system

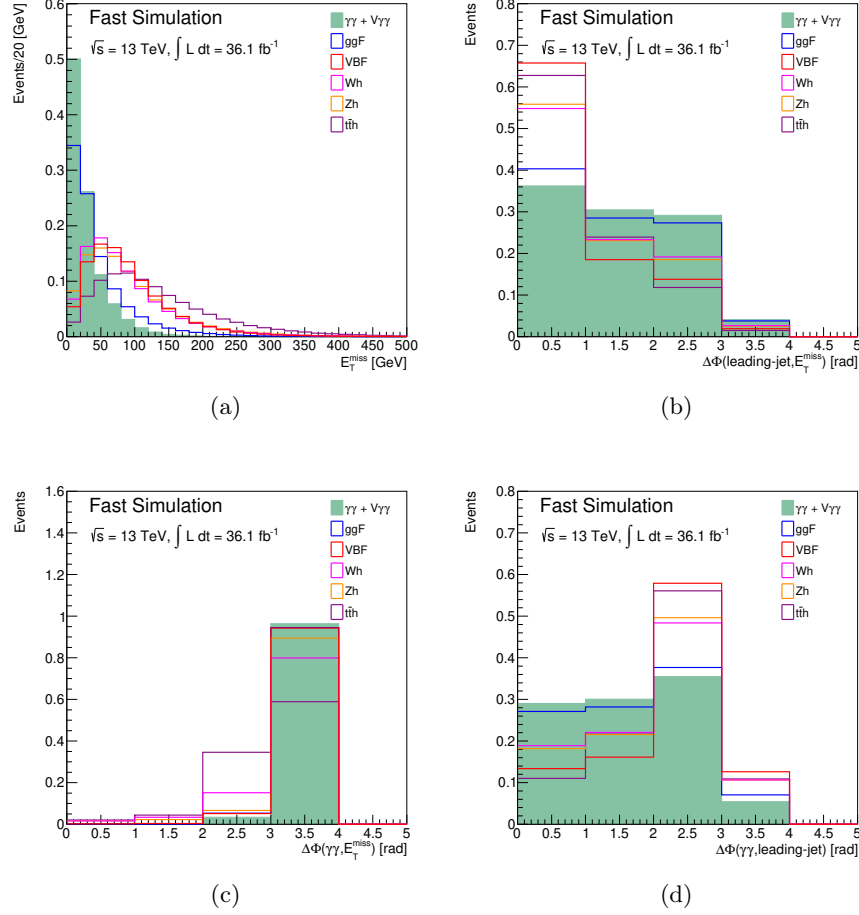


Figure 4.6: Signal and background distributions of: (a) reconstructed missing transverse energy, (b) Azimuthal angle separation between the leading jets and the missing transverse energy, (c) Azimuthal angle separation between the di-photon system and the missing transverse energy (d) Azimuthal angle separation between the di-photon system and the leading jets.

4.4 The DNN configuration

The architecture of the DNN was configured using the ROOT [194] package and the TMVA data processing system [195]. The structure of the DNN implementation is as follows: an input layer with 13 units that represent kinematic features as explained in the previous section, three hidden layers with 26, 26 and 13 units, and a single output layer. Figure 4.7 shows a schematic illustration of this structure. The cross-entropy (see equation 4.2) loss function is used in this DNN to classify signal events from background events. This is combined with zero-mean Gaussian distributed random variables and adjusted initialization weights [196].

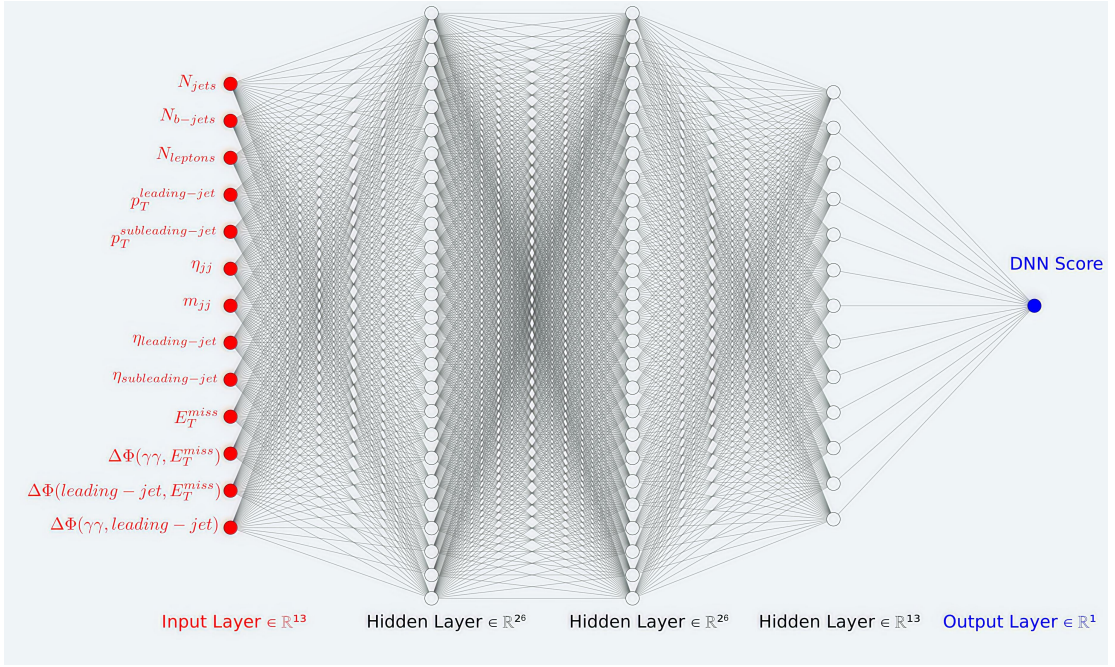


Figure 4.7: The DNN architecture configuration for the classification of events as background or signal in this study.

The activation functions of neurons and ANNs, in general, are extremely important. They're used to compute the weighted sum of the inputs. This is done for each neuron, which is a simple computational element that helps the ANN learn the mappings between input and output data. Also, these functions are used to normalize the output of each neuron so that it is between 0 and 1 or -1 and 1. Linear and nonlinear activation functions are the two types of neuron response functions, with nonlinear activation functions being the most commonly used. Sigmoid, Rectified Linear Unit (ReLU), tanh, and sigmoid are just a few examples of nonlinear activation functions. The activation function used has a significant impact on the training dynamics and performance of the model [197]. However, since these activation functions are computed hundreds or thousands of times, especially in the case of DNNs, they must be computationally efficient.

A sigmoid function and hyperbolic tangent functions were used as activation functions. These were used for both the hidden layer and the output layer. The two activation functions have a few similar properties, for example, they are sigmoidal functions and have a vanishing gradient problem. The output neuron, on the other hand, is normalized and lies between 0 and 1 or -1 and 1. The sigmoid is mathematically represented by the equation 4.6, while the hyperbolic tangent is represented by the equation 4.7. When applied to DNNs, the sigmoid function faces the vanishing gradient problem, which leads to slow convergence of the optimization [198].

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4.6)$$

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4.7)$$

4.4.1 Backpropagation

Backpropagation was used to train the DNN, a technique widely used for fully supervised learning of ANNs [199, 200]. This algorithm uses gradient descent and takes its name from the ‘backward propagation of errors’ during the training phase of ANNs. Since the gradient is computed backwards through the network, the first computation occurs at the weights of the last layers, while the weights of the first layers are computed last. All this happens in reverse through the intermediate hidden layers, hence the name. So these steps come after forward propagation, where the output of the network is computed from the input layer to the output layer. The sequence of these steps is performed iteratively for each training pattern, adjusting the weights until the error between the predicted output and the actual output is significantly small [201].

However, this approach has a slow convergence problem [202–204]. Implementing an extremely low learning rate [205–209] is one technique to address this problem. Several studies have addressed the question of how to improve the convergence time of the backpropagation algorithm. Heuristic approaches, flat-spots, and Quasi-Newton were used to speed up the training time of the backpropagation algorithm [210]:

$$\Delta w_{ij}(n) = -\alpha \frac{\partial E}{\partial w_{ij}} + \mu \Delta w_{ij}(n-1), \quad (4.8)$$

where α is the learning rate and μ is the momentum. This equation gives the n^{th} correction step. The partial derivative $\frac{\partial E}{\partial w_{ij}}$ is calculated as:

$$\frac{\partial E}{\partial w_{ij}} = -\delta_j(n) o_i(n-1), \quad (4.9)$$

where $\delta_j o_i$ is the propagated error up to neuron j and o_i is the output of the i^{th} neuron in the forward direction [211]. The primary goal of this algorithm is to reduce the partial derivative of the cost function with respect to the network weights [212]. The approach works by minimizing the partial derivative of the error (cost) function E in terms of network weights, i.e., $\partial E / \partial w$.

4.5 Fully Supervised Learning Classification

The fully supervised learning technique was applied to each separate SM Higgs boson production mechanism and the entire background was examined for this study. The signal was labeled 1 and the background was labeled 0. Figures 4.8 and 4.9 show the DNN response distributions for ggF with VBF and the associated productions, respectively. This has shown a clear separation of signal and background processes. By comparing the results of the training and test samples, the above figures also determine

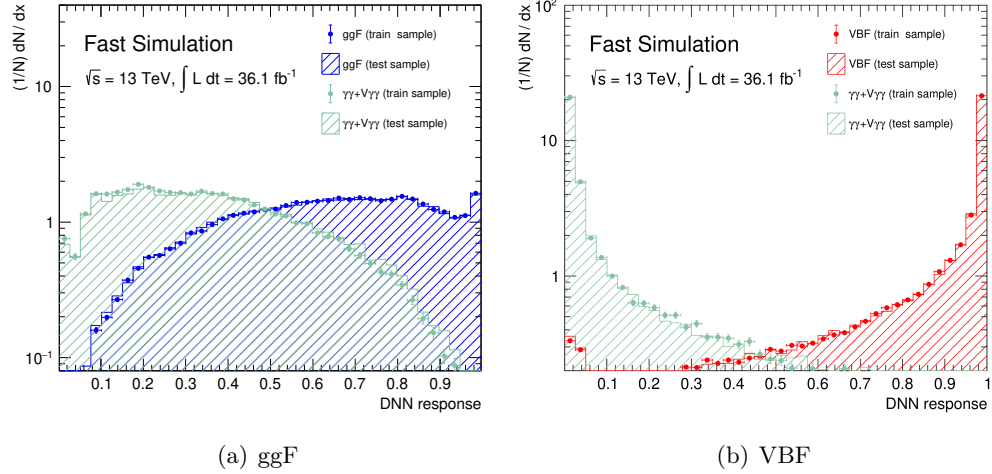


Figure 4.8: Fully supervised learning DNN response distributions for the Higgs production, by: (a) gluon-gluon fusion and (b) VBF.

whether the model is overtrained. In the best case, the DNN response distribution functions for training and testing samples must be similar to prove that the classifiers are not overtrained. This study does take into account any experimental or theoretical uncertainties that could lead to a slight decrease in performance. The Receiver Operator Characteristic (ROC) curve is used to evaluate the separation performance of the DNN classifier. This can be used to quantify exactly how well the classifier can distinguish between signal and noise. The performance of the model is measured by the area under the curve (AUC), which is calculated by measuring the total two-dimensional area under the ROC curve. Figure 4.10 shows the background rejection versus signal efficiency for several Higgs production processes, including ggF, VBF, and associated productions ($t\bar{t}h$, Zh and Wh). Each of the production mechanisms has a different separation power depending on the event configuration. The ggF production mechanism has an AUC of 78%, which could be due to the fact that it populates a phase-space comparable to the background. Due to the variety of signatures and topologies, the results for the $t\bar{t}h$ stand out, although the VBF and associated productions are comparable and better than ggF.

4.6 Semi-supervised learning classification

4.6.1 Supervised classification with unlabelled signals

The fully supervised classification strategy with a mixture of signals¹ (as described in Section 4.2.4.2) is tested for its ability to distinguish between different SM Higgs boson production processes. In this method, five separate production processes are combined to produce a mixture of signal samples. As a result, this mixture is labeled 1, while the

¹The mixture is referred to as unlabelled in the context of this approach since it is comprised of different production mechanisms.

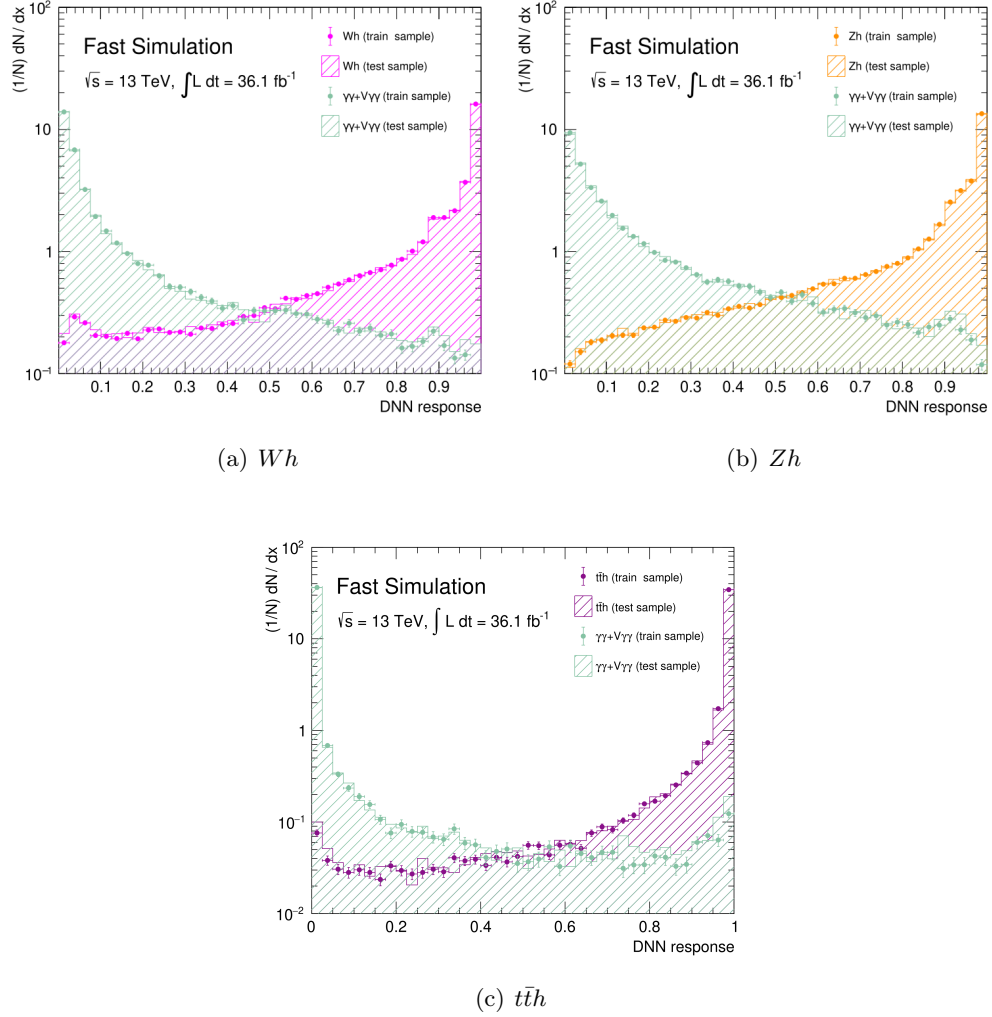


Figure 4.9: Fully supervised learning DNN response distributions for the SM Higgs boson associated production, with: (a) W vector boson, (b) Z vector boson or (c) a top quark pair.

background is labeled 0. Figure 4.11 shows the DNN classifier response distribution of the individual Higgs productions and the total background. The performance of this strategy is measured by the ROC curves, which show that the separation performance of the fully supervised learning method with unlabelled signal processes is lower than that of the fully supervised learning.

However, since the ggF is the major mechanism responsible for more than 85% of the Higgs generation, its performance has remained stable because the mix is proportional to the cross sections. This observation is explained by the different production mechanisms having different event yields. Therefore, the predicted cross-sections are increased to quantify the relative contribution of each production mechanism to the classifier. The

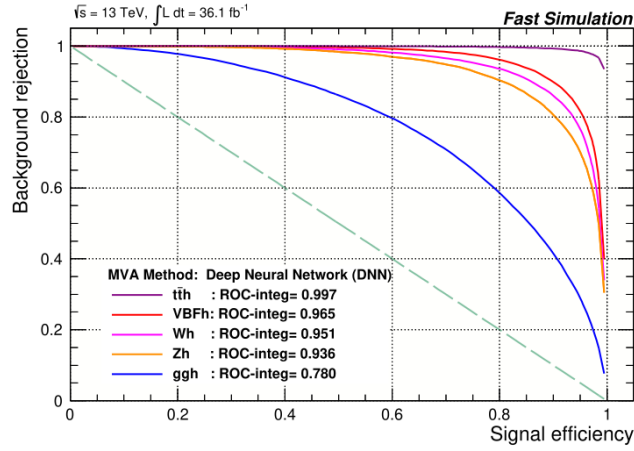


Figure 4.10: ROC curves of DNN classification with fully supervised learning for different Higgs production mechanisms.

additional production mechanisms, VBF, Wh , Zh , and $t\bar{t}h$ are normalized to different proportions of the predicted ggF production yield.

4.6.2 Effect of signal event yield on the event classification

Unlabelled supervision is used to test the impact of the event yield of different production mechanisms on the DNN classifier. All production mechanisms are grouped under the same label and have an expected cross-section of 36.1 fb^{-1} . The predicted cross-sections of VBF, Wh , Zh , and $t\bar{t}h$ were consistently improved by using the ggF method. This improvement was achieved by assigning specific fractions to each process, 0.25, 0.5, 0.75, and 1, respectively, indicating the fraction of the cross-section attributable to the ggF method. The ROC curves for each of the production techniques are shown in Figure 4.12, which compares the effects of increasing the ggF cross-section on signal efficiency.

The results show that the efficiency of an unlabelled classifier is determined by the relative contribution of each mechanism as well as the complexity of its topological features. Both the contributions of the production mechanisms and the complexity of the topological features affect the efficiency of the classifier with weakly supervised learning. The results from Section 4.6.1 are summarized in Table 4.2, where the integral of the ROC curves for all production processes increases with increasing ggF cross-section fraction.

4.6.3 Weakly supervised learning classification

The study of Higgs production mechanisms with weakly supervised learning involves mixing signals and background processes in the di-photon invariant signal mass window from 120 to 130 GeV. Outside the signal mass window is the region of the sideband that lies between 105 and 160 GeV. This region is made up entirely of pure background, while the mass window is made up of a combination of signal and background from this

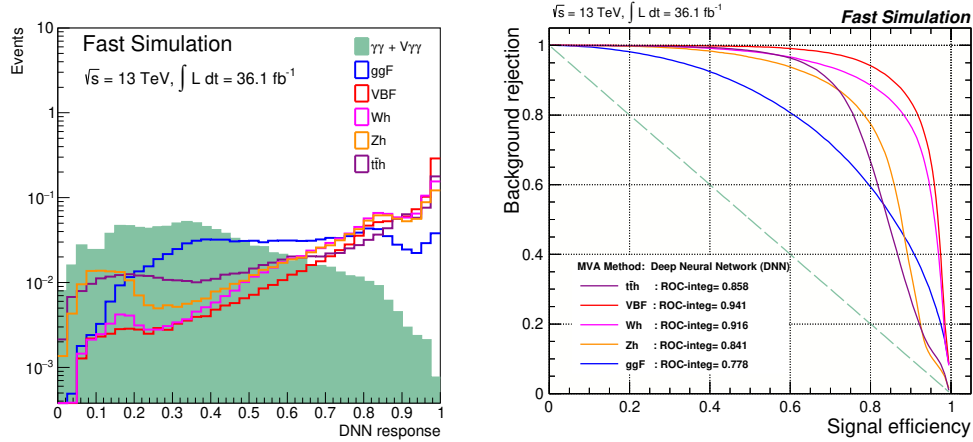


Figure 4.11: Results of fully supervised learning with unlabelled signals. Left: DNN response distributions. Right: ROC curves, for different Higgs production modes.

Process	$\sigma_{process}$	ROC integral normalised to			
		0.25 σ_{ggF}	0.5 σ_{ggF}	0.75 σ_{ggF}	σ_{ggF}
ggF	0.778	-	-	-	-
VBF	0.941	0.946	0.951	0.954	0.961
Wh	0.916	0.934	0.941	0.943	0.944
Zh	0.841	0.892	0.916	0.920	0.920
tth	0.858	0.975	0.984	0.986	0.987

Table 4.2: Relative contribution of each signal to the performance of the DNN classifier, for fully supervised learning with unlabelled signal (Section 4.6.1). Signals are normalized to different proportions of the expected events from the ggF production mechanism.

region. The purpose of this study is to find out how well the DNN can distinguish the signal from the signal mass window of an unknown mixture of signal and noise. This is done with respect to the pure background in the sideband.

The ROC AUCs in Figure 4.13 show that the weakly supervised learning strategy is able to discriminate between different signal candidates. However, the results of both fully supervised learning and supervised learning with unlabelled signals are not comparable. Table 4.3 summarises the overall results of the three approaches. The complexity of the phase-space investigated in this study can be related to the performance of the weakly supervised learning strategy. As a result, the guided weakly supervised learning with topological constraints was introduced, which is studied in detail in Section 4.6.4.

4.6.4 Weakly supervised learning with topological requirements

Using the classification of SM Higgs boson production modes, the performance of the different machine learning algorithms investigated in this study was evaluated and com-

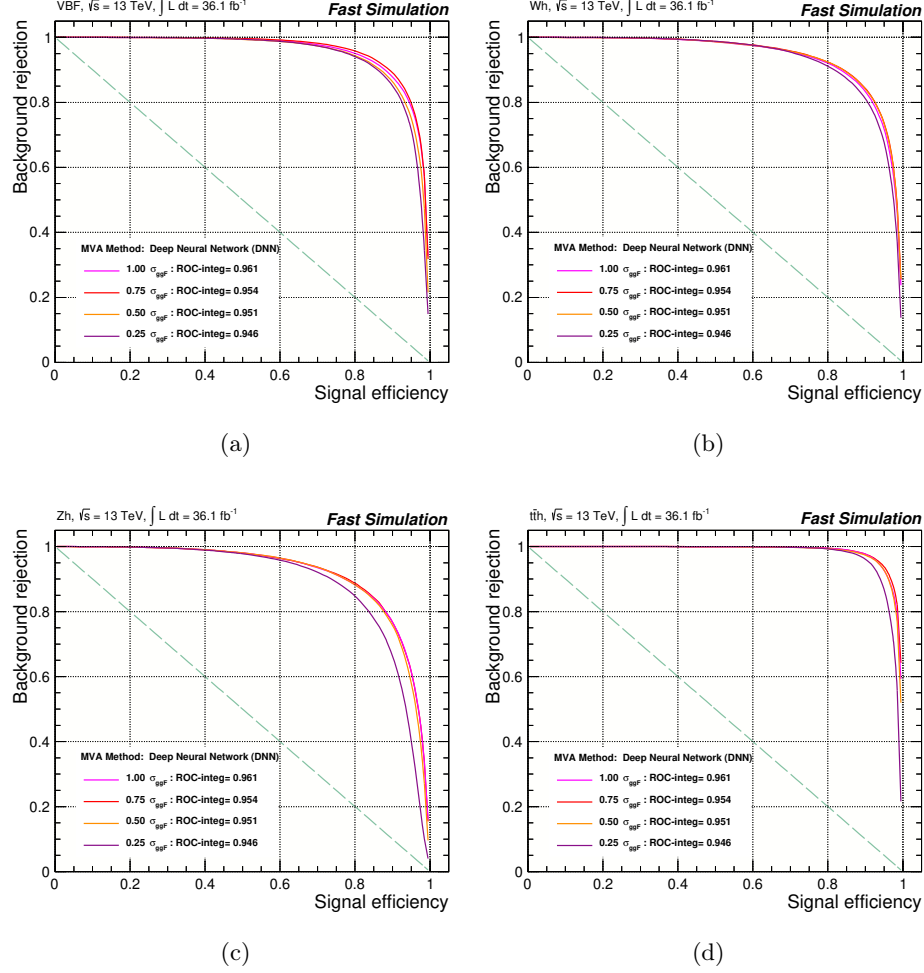


Figure 4.12: ROC curves from fully supervised learning with unlabelled signals, (a) VBF, (b) Wh , (c) Zh and (d) $t\bar{t}h$ are normalized to different proportions of the expected ggF production mechanism events.

pared. Compared to fully supervised learning with unlabelled signal and weakly supervised learning, the results show that fully supervised learning gives better results. However, the performance is comparable to that of supervised learning with unlabelled signals. The differences between these methods are due to the relative contributions of each SM Higgs boson production mechanism. A ROC integral difference of 0.2% is seen in the case of ggF, which has the highest effective cross-section at the LHC and approximately the same topology as the main background (see Table. 4.3). The ROC integral difference is 14% since the $t\bar{t}h$ production mechanism contributes less than 1% of the entire production cross-section. As a result, the weakly supervised learning strategy yields the worst results, and a different approach called guided weakly supervised learning was developed to address this problem. This method restricts training to topological signatures produced from specific production processes and is then evaluated and

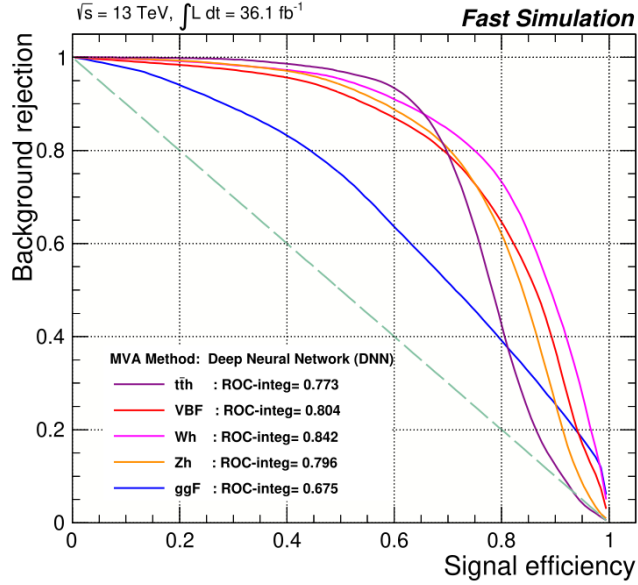


Figure 4.13: ROC curves from weakly supervised learning using the DNN classifier for different Higgs production mechanisms. Signals and background are normalized to their cross-sections.

verified using various SM Higgs boson generation methods. In contrast to the ggF and $t\bar{t}h$ production mechanisms, the VBF production mechanism involves the scattering of two quarks, resulting in two hard jets in the forward and backward regions of the ATLAS detector with a high invariant di-jet mass. In this context, the topology of the VBF is defined as occurrences with at least two jets, which is followed by a signal significance scan using m_{jj} and the pseudorapidity gap between the two leading jets. The signal significance scan is also used to increase the relative contribution of the VBF production mechanism while reducing the background yield as much as possible. Before using VBF selection, Figure 4.14 compares the ROC performance of the three approaches.

The results for the weakly supervised learning algorithms with topological requirements are shown by the highlighted dashed lines. The DNN is trained in four different regions, each specified by the significance of the signal. Weakly supervised learning with topological constraints has led to a variety of results. These results support the ability of the technique to distinguish VBF from other production processes. Therefore, the method is comparable to fully supervised learning. The associated production of the Higgs boson is also tested with a W gauge boson, similar to the VBF test. In this case, two topological features from the Wh final states are used to train the DNNs individually, which includes both leptonic and hadronic decays of the W boson. Figure 4.15 shows the results of fully supervised learning combined with weakly supervised learning and supervised learning with unlabelled signals based on this test. The performance of the weakly supervised learning appears to improve with the topological constraints, suggesting that it is able to isolate the Wh process. Higgs boson production associated with top-antitop quarks

Process	Full Supervision	Unlabelled Full Supervision	Weak Supervision
ggF	0.780	0.778	0.675
VBF	0.965	0.941	0.804
Wh	0.951	0.934	0.842
Zh	0.936	0.841	0.796
$t\bar{t}h$	0.997	0.858	0.773

Table 4.3: Comparison of the integral of ROC curve AUCs from different learning methods.

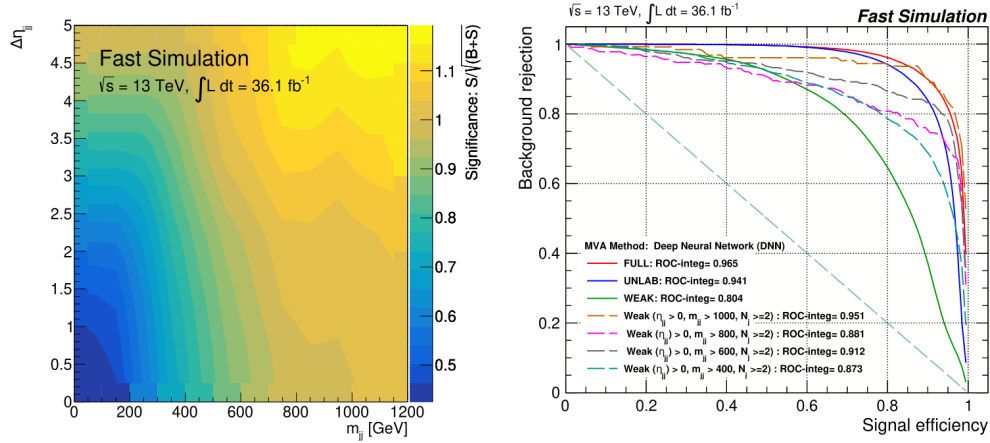


Figure 4.14: Weakly supervised learning results with VBF topological selections. For the VBF production mode, the diagram on the left shows a two-dimensional scan of the signal significance with respect to m_{jj} and $\Delta\eta_{jj}$. The expected signal significance is represented by the z axis. The ROC curves for the VBF production mechanism with various learning methods are shown in the graph on the right (see text). Weakly supervised learning results are not as good as unlabelled or fully supervised at background rejection.

is also studied using the proposed guided weakly supervised learning method. Given that top-antitop quarks account for less than 1% of the total Higgs boson production. As a result, it becomes difficult to extract the signal process. In the weakly supervised learning situation, the same finding was made, but in this case, the test is performed by requiring two topological characteristics.

The total hadronic and semi-leptonic decays of the top quark from the $t\bar{t}h$ process are represented by these topological characteristics. Here, at least two central jets provide for the hadronic decay and at least one b-jet and one lepton for the semi-leptonic decay. In terms of ROC curves, Figure 4.16 shows the performance of guided weakly supervised learning for the two topologies. Integral values from ROC, which quantify the methodology's ability to extract tiny signals, are used to compare their performance. Figure 4.16 shows the ROC curves for the guided weakly supervised learning for the two typolo-

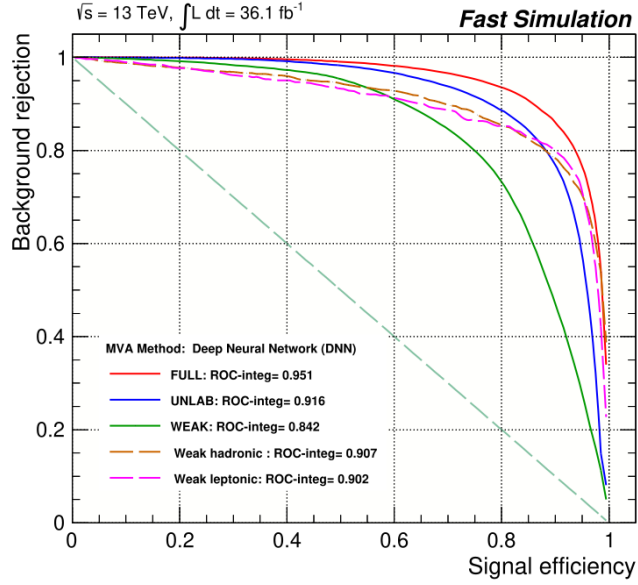


Figure 4.15: The solid lines show the ROC curves in inclusive Wh final states using different learning approaches. The leptonic and hadronic Wh topology requirements with weakly supervised learning are shown with dashed lines.

gies. The ROC integral values are a representation of the method's ability to extract small signals. Table 4.4 compares the efficiency of inclusive weakly supervised learning with that of guided weakly supervised learning for the VBF, Wh , and $t\bar{t}h$ production mechanisms.

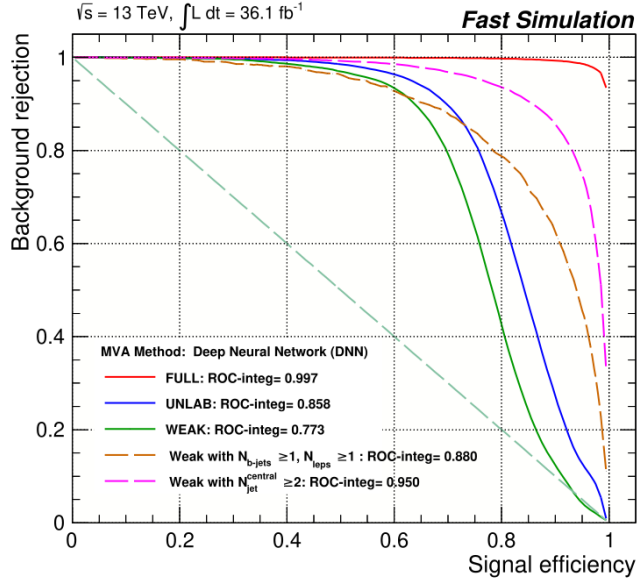


Figure 4.16: The solid lines show the ROC curves for a range of different learning methods in inclusive $t\bar{t}h$ final states. The dashed lines show the weakly supervised learning with $t\bar{t}h$ topological requirements.

Process	Weak Supervision	ROC Integral
VBF	Inclusive	0.804
	$N_{jets} \geq 2, \Delta\eta_{jj} > 0, m_{jj} > 1000 \text{ GeV}$	0.951
	$N_{jets} \geq 2, \Delta\eta_{jj} > 0, m_{jj} > 800 \text{ GeV}$	0.881
	$N_{jets} \geq 2, \Delta\eta_{jj} > 0, m_{jj} > 600 \text{ GeV}$	0.912
	$N_{jets} \geq 2, \Delta\eta_{jj} > 0, m_{jj} > 400 \text{ GeV}$	0.873
Wh	Inclusive	0.842
	Hadronic	0.907
	Leptonic	0.902
$t\bar{t}h$	Inclusive	0.773
	$N_{b-jets} \geq 1, N_{leptons} \geq 1$	0.880
	$N_{b-jets} \geq 2$	0.950

Table 4.4: VBF, Wh and $t\bar{t}h$ ROC integrals with different topological requirements, compared to inclusive weakly supervised learning.

4.7 Discussion and conclusions

The presence of massive SM backgrounds hinders the search for new resonances at the EW scale. Moreover, neither the inclusive nor the model-dependent searches were successful in identifying corners of the phase-space with potential excesses over the SM-only

hypothesis. This may suggest that novel resonances exist in delicate topologies that are difficult to detect using standard and popular search methods. The use of ML at the LHC could be crucial for a more in-depth investigation of the available phase space. In this study, we analyze the performance of different ML approaches using the SM Higgs boson with di-photon decay.

The study helps to understand the performance of ML algorithms in situations where the particle is produced with different production mechanisms. Compared to fully supervised learning, the performance of supervised learning with an unlabelled signal was better for a signal sample where the different production mechanisms were unlabelled. Weakly supervised learning with an unlabelled signal sample is also compared with these two approaches. Compared to full supervised learning, the latter performs significantly worse for all production mechanisms considered (see Table 4.3).

The $t\bar{t}h$ mechanism, which produces the largest number of different signatures, has the largest impact. This is followed by the VBF mechanism, which gives a different topology than the second one. This leads to the realization that in the presence of signatures and topologies, weakly supervised learning alone is not always effective in distinguishing the signal from the background. It is worth noting that hyperparameter optimizations were performed to ensure that the DNNs perform to their full potential.

Prior to the introduction of weakly supervised learning, signatures and loose topological constraints are introduced to counteract this effect. For the purpose of this study, this strategy is referred to as guided weakly supervised learning. This is because the approach evaluates the impact on signal efficiency and background rejection for different production mechanisms, showing significant gains when weakly supervised learning is implemented. Weakly supervised learning, as used in this study, has the advantage that it does not require the use of a signal model. As a result, it is necessary to control the phase-space in which the sideband and the signal region are confronted with each other. The search is not affected by the phenomenology of a model with a particular set of parameters, since the signature and topology requirements are dictated by physics considerations.

Chapter 5

Application of Semi-supervision in the search for resonances with the $Z\gamma$ final state at the LHC

5.1 Introduction

This study focuses on the search for heavy resonances decaying to $Z\gamma$ final states, where $Z \rightarrow e^+e^-$ or $Z \rightarrow \mu^+\mu^-$. The data used in this analysis correspond to the integrated luminosity of 139 fb^{-1} . In addition, the research continues with the primary goal of laying the groundwork for the search for heavy resonances in the electroweak scale. Therefore, it is performed with topological requirements using ML techniques. Unlike fully supervised learning, which is known to assume full knowledge of the underlying model, the semi-supervised approach, especially weakly supervised learning, allows the extraction of new information from data with only partial knowledge. Therefore, this approach is well suited for the proposed analysis, as it offers the advantage of using sidebands or control regions that have minimal dependence on simulations when modeling background. In addition, we evaluate the performance of weakly and fully supervised learning in the inclusive and exclusive regions of phase-space, tailored for different production mechanisms of the signals at the LHC.

The application of ML techniques in the search for resonances can play a crucial role. These techniques can help in finding resonances that may be missed in the corners of the phase-space that are either unexplored or poorly covered by the search. The fully supervised learning paradigm has proven to be an effective tool that can be used to explore the corners of the phase-space of interest. This is also true for the phase-space where the Higgs boson SM has been observed. However, as effective as this tool may be, it is still hampered if the new physics populates the phase-space differently from than model used as a baseline. This is where the implementation of weakly supervised learning comes in, which has the advantage of detecting topological differences between the potential signal region and the sidebands. Weakly supervised learning is restricted to the topological configurations due to the complexity of the phase-space under study and the way the background is populated. These configurations of the phase-space include

a number of b -tagged jets, a number of leptons, vector boson fusion topology, etc.

The focus of this research is to search for resonances with $H \rightarrow Z\gamma$ in the final states in predefined categories shown in Table 5.1 where H is a Higgs-like scalar. The purpose of this study is to scan the 130 - 600 GeV range of the invariant mass of the $Z\gamma$ to verify the capability of the proposed methodology, which will then be used to prepare a search for new phenomena at high masses in $Z\gamma$ final states at the LHC. Since this is an ongoing study within the Wits ICPP group, this work focuses more on the 150, 200, 300, and 500 GeV mass points.

5.2 Data and Simulated Samples

5.2.1 Data

The analysis uses pp collision data with a bunch crossing interval of 25 ns collected in 2015-2018 at $\sqrt{s} = 13$ TeV. The average number of inelastic interactions per bunch crossing during this data collection period was 34 and varied from 24 in the 2015-2016 data to 37 in the 2017-2018 data. Events recorded under stable beam conditions, where the trigger system, tracking devices, and calorimeters were operating, and meeting the data quality criteria are considered. The data sample is applied according to data quality requirements to ensure good working conditions for all detector components, corresponding to an integrated luminosity of $139.0 \pm 2.4 \text{ fb}^{-1}$.

5.2.2 Backgrounds

The dominant background of this analysis is the non-resonant $Z\gamma$ production generated using MC. The $Z + jets$ background is estimated from the data by reversing the photon identification criteria (ID). The MC samples are used only to normalize the categories in the optimization to choose the background fit functions. The normalization of the $m_{Z\gamma}$ background and the shape of the $Z + jets$ are obtained directly from the data, using the purity study. In the statistical fit, the shape and $m_{Z\gamma}$ spectrum are obtained directly from the data.

5.3 Object Reconstruction and Overlap Removal

Photons, leptons (electrons and muons), and jets are used in the selection of this analysis. Their reconstruction techniques are summarised in Chapter 3.4. To avoid unnecessary double-counting, objects which pass selection requirements for more than one category are removed. An example of this is the electrons which may be identified as jets or a photon reconstructed as an electron. This is achieved by the process termed overlap removal which resolves the ambiguity. For the purpose of this analysis, the following objects are removed to maximize signal efficiency:

- Electrons that lie within $\Delta R = 0.4$ of any photon.
- Muons that lie within $\Delta R = 0.4$ of any photon.

Mass Point (GeV)	Process	Category	Sideband Region	Signal Region	Signal Injection
150	Zh or Wh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	3147	1664	81
	Zh or Wh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	1875	1032	64
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	1103	603	49
200	ggF	Inclusive	13175	6444	160
	Zh or Wh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	808	394	40
	Zh or Wh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	839	414	40
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	498	245	30
300	ggF	Inclusive	4173	2022	90
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	237	113	21
500	ggF	Inclusive	704	343	37
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	56	29	11

Table 5.1: Events yield for the considered processes and signal injection rates for VBF, ggF, Zh and Wh signal samples at different mass points.

- Jets that lie within $\Delta R = 0.4$ of any photon.
- Jets that lie within $\Delta R = 0.2$ of any electron.
- Electrons that lie within $0.4 \Delta R$ of any jet.
- Muons that lie within $\Delta R = 0.4$ of any jet.

5.4 Data and Kinematic Features Selection

Table 5.1 shows the number of expected event yields for both the sideband and signal region for all predefined categories for different mass points considered for this study. The width of the invariant mass $Z\gamma$ for the sideband region is defined as 12% of the center mass of the resonance Higgs-like signal, while the signal region is 6%. Different signal production mechanisms (Wh , Zh , ggF and VBF) were injected. These injected numbers are defined by 2σ , where σ is the statistical uncertainty of the background in the signal region, given by:

$$\sigma = \sqrt{B_{MW}}, \quad (5.1)$$

where B_{MW} is the number of background events in the mass window region. The signal region for the 200 GeV mass is defined as $194 < m_{\ell\ell\gamma} < 206 \text{ GeV}$ whereas the sideband is between $182 < m_{\ell\ell\gamma} < 194 \text{ GeV}$ and $206 < m_{\ell\ell\gamma} < 218 \text{ GeV}$. In this study, the background events in the sideband region are represented by sample 1, while sample 2 consists of the background and signal in the mass window region. This serves the purpose of the weakly supervised learning technique. Figure 5.1 shows the invariant mass with an injection of signal for the categories covered at 200 GeV mass point.

Feature engineering plays an important role in the development of ML models. This usually involves approaches such as feature extraction and feature selection. Feature extraction is about reducing the dimensionality of the data to make it easily manageable for the model. This can be done with algorithms such as principal component analysis (PCA) and edge detection. Feature selection, on the other hand, is about determining

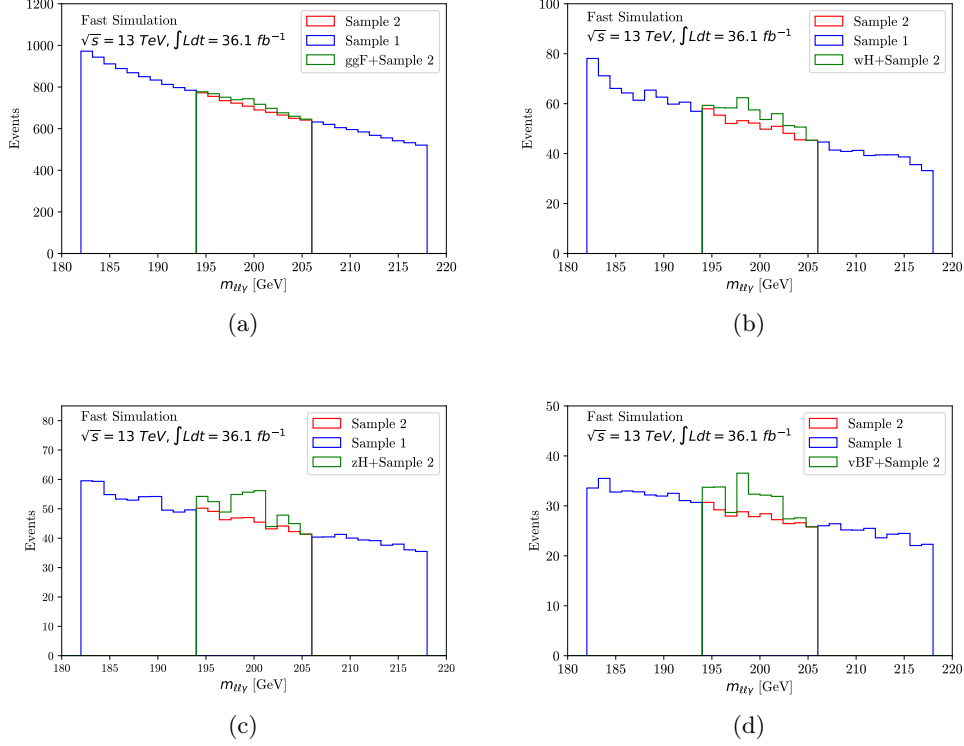


Figure 5.1: Invariant mass with an injection of signal for the following categories: (a) Inclusive, (b) $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$, (c) $N_{jets} \geq 2$; $60 < m_{jj} < 120 \text{ GeV}$ and (d) $N_{jets} \geq 2$; $\Delta\eta_{jj} > 2$; $m_{jj} > 300 \text{ GeV}$.

which features are important, redundant, and irrelevant. This is done by ranking features using specific algorithms that help determine which features to remove and which to prioritize for modeling. The correlation matrix is one of the techniques for this purpose, as it indicates the relationship between the dependent and independent features.

The main purpose of this step is to select the most important features in the data. The number of selected features gives information about the number of input units used in the DNN configuration. It can also reduce the computation time of the model as it is trained on less important features. This is achieved by removing redundant features that could add unnecessary complexity. The correlation matrix of kinetic features for this particular study are all shown in Figures 5.2 and 5.3. This is specifically for the category $N_{jets} \geq 2$; $\Delta\eta_{jj} > 2$; $m_{jj} > 300 \text{ GeV}$ for the mass point 200 GeV for both the background and the signal. The correlation between two pairs of features is indicated by a value between -1 and 1 , where 0 indicates no correlation between the two variables. Both 1 and -1 indicate strong positive and strongly negative correlations, respectively.

For example, the Figure 5.2 indicates that there is no significant correlation between m_{jj} and variables such as η_{j1} , η_{j2} , ϕ_{ET}^{miss} and $\Delta\phi(Z\gamma, E_T^{miss})$. The correlation between

these variables and m_{jj} ranges from -0.5% and 0.6% . Together with other correlations below 20%, this means that there is less redundancy of important information, which is

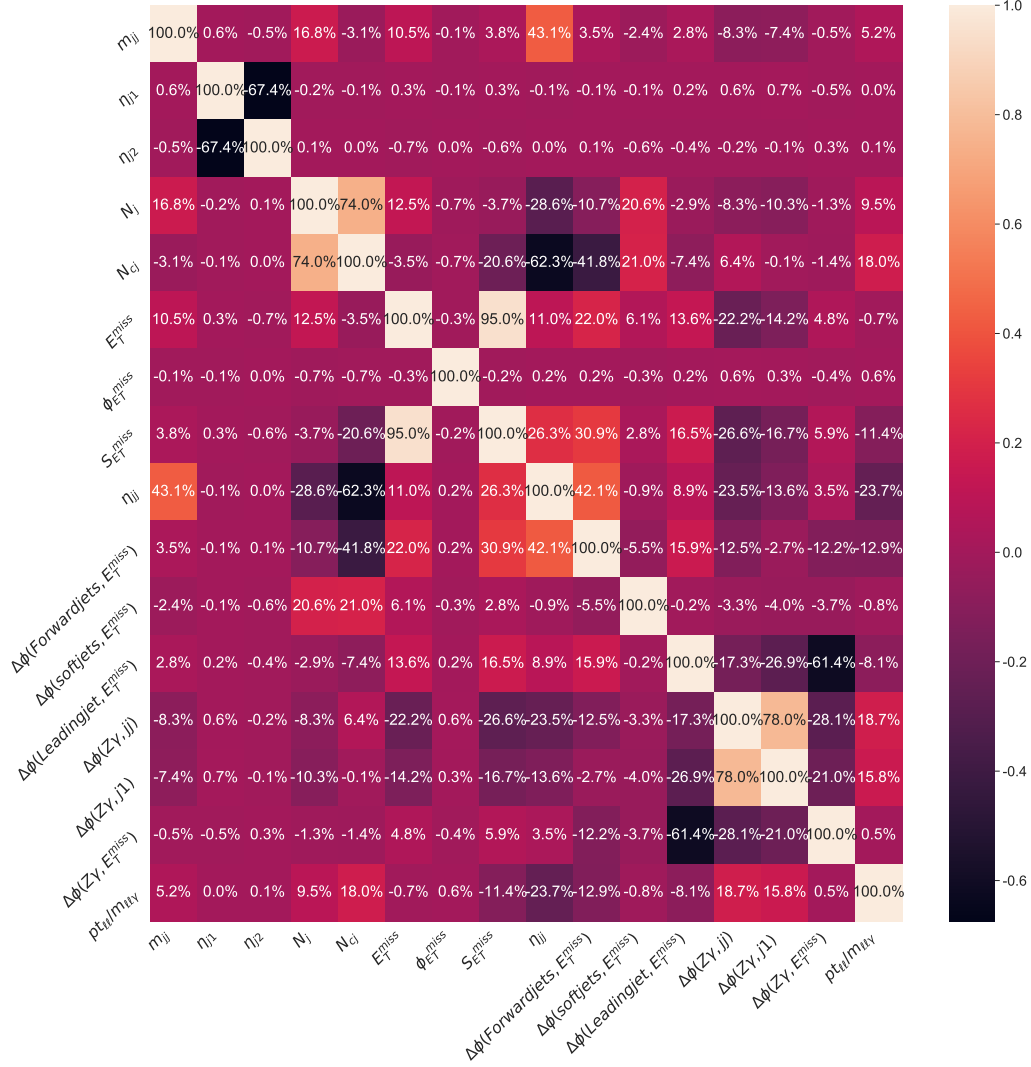


Figure 5.2: Correlation matrix of the implemented features/kinematics for the 200 GeV background VBF category.

not the case for highly correlated pairs of features. There appears to be a high correlation between η_{j1} and η_{j2} , due to some degree of similarity between the two features, both for signal and background. Overall, there is no significant correlation between the kinematic features. This is also evident in the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category shown in Appendix A, Figures A.5 and A.6. The effect is somewhat different, however, as there is a noticeable correlation for both the inclusive and the $N_{jets} \geq 2$; $60 < m_{jj} < 120 \text{ GeV}$

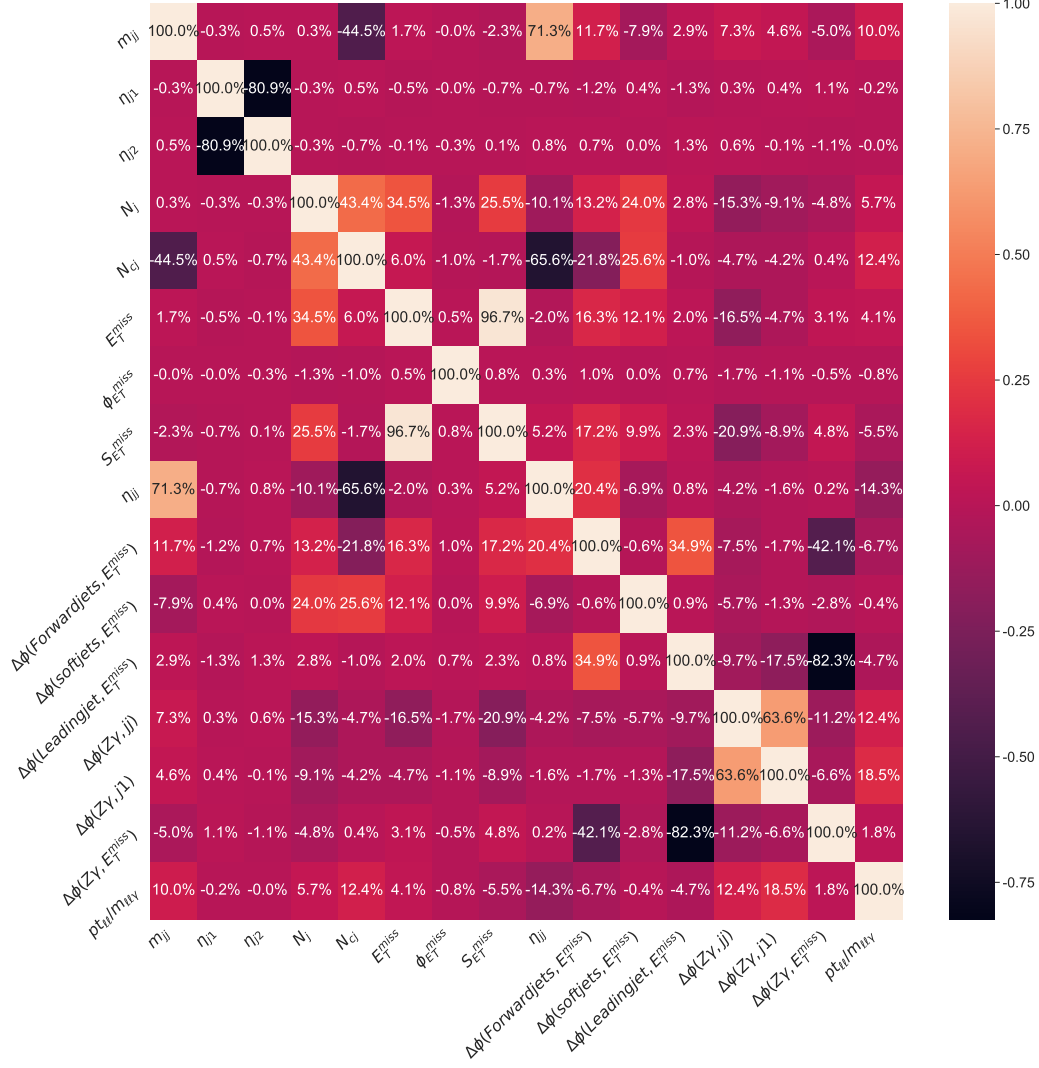


Figure 5.3: Correlation matrix of the implemented features/kinematics for the VBF signal for 200 GeV.

categories, which showed variation in correlations between some pairs of features (see Figures A.1 - A.4 in Appendix A). The information in the correlation matrix makes it valuable for training the ML algorithm. These selected features described below were chosen and their distributions can be seen in Figure 5.3. This is demonstrated for the VBF category, as the same set of features is used for the other categories.

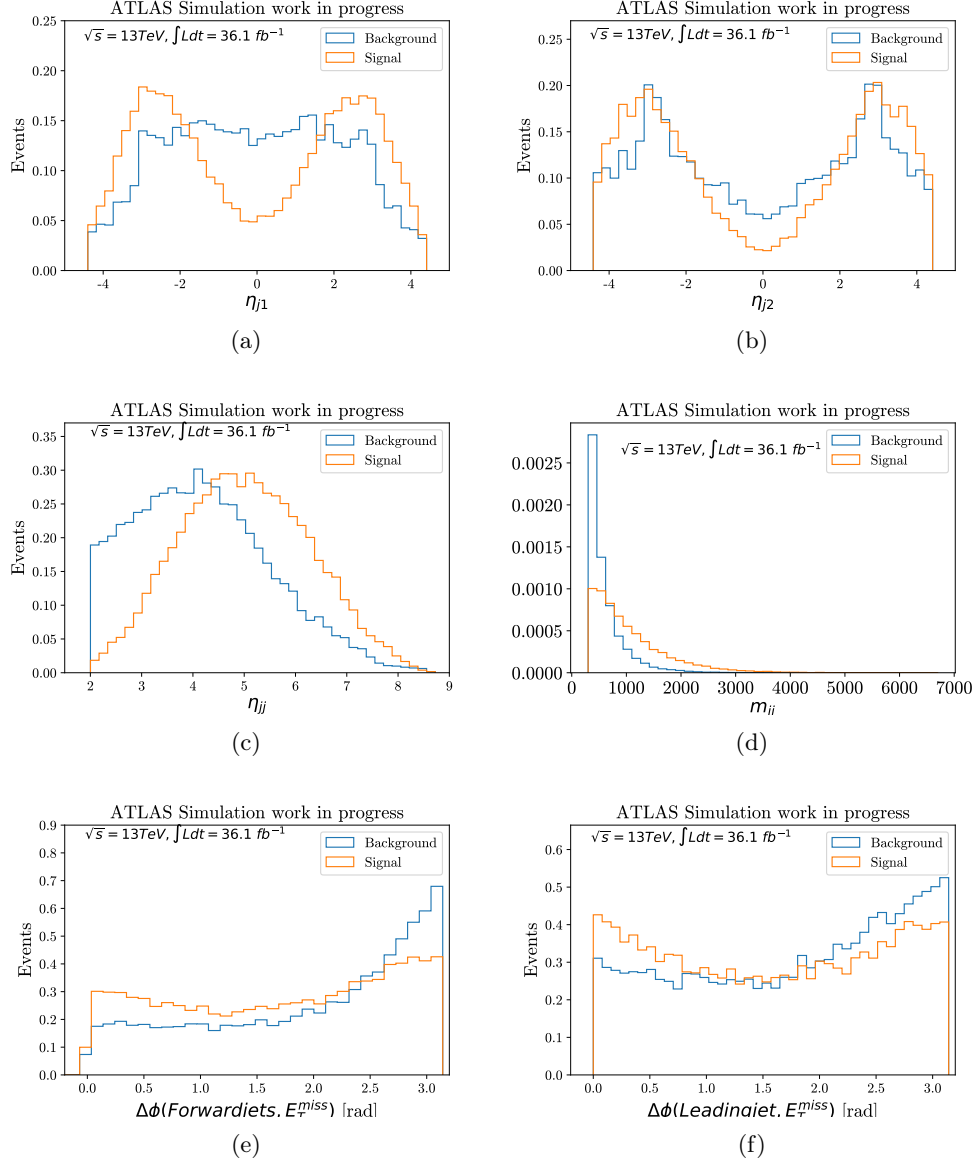


Figure 5.3: The signal and background distribution of: (a) pseudorapidity of the leading jet, (b) pseudorapidity of the sub-leading jet, (c) pseudorapidity of the dijet system, (d) invariant dimass jet of the leading and sub-leading system, (e) Azimuthal angular separation between the forwards jets and missing transverse energy and (f) Azimuthal angular separation between the leading jets and missing transverse energy.

5.4.1 Data Processing

Similar to feature engineering, data processing plays a crucial role in ML modeling [213–215]. It ensures that the raw data collected is transformed into meaningful data from which the model can easily make sense of. In many cases, the collected raw data

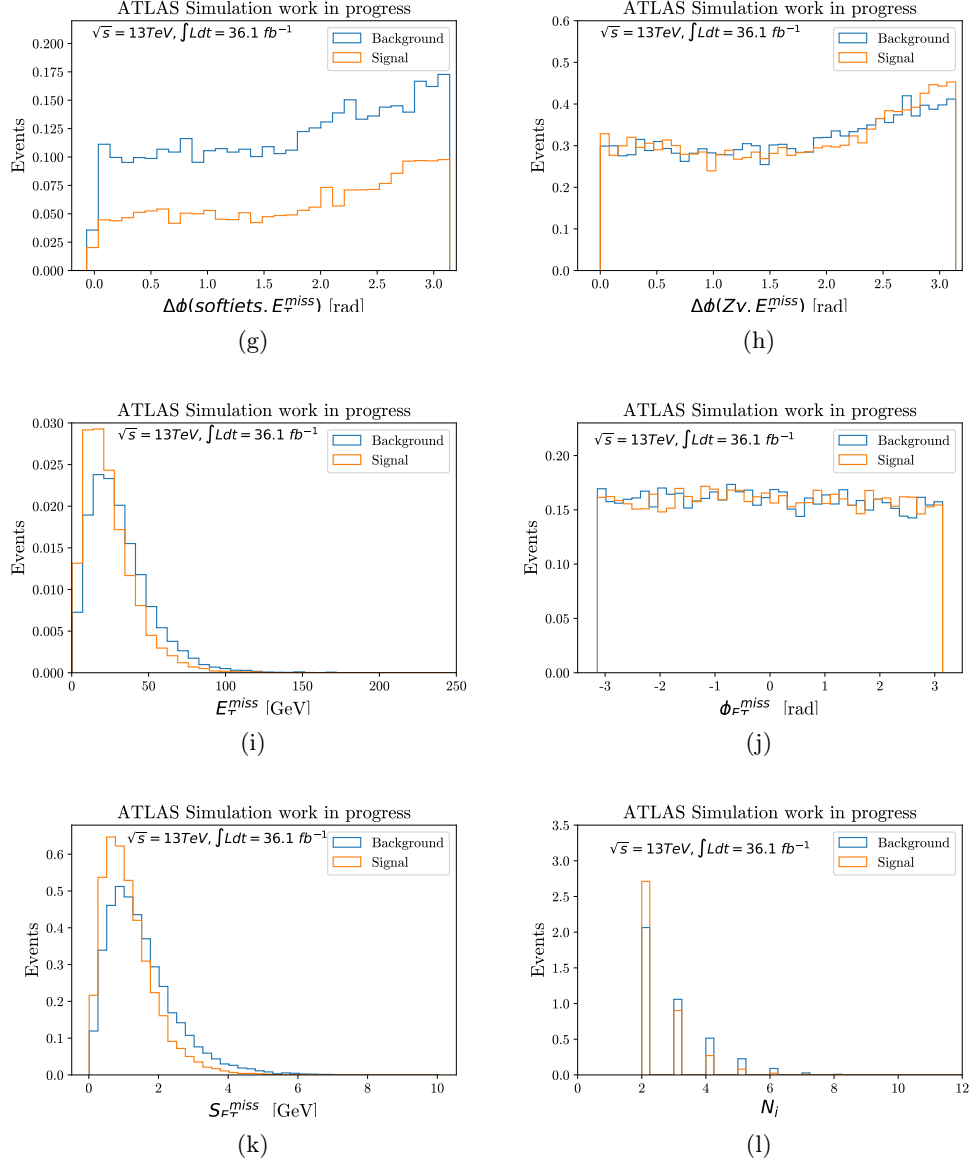


Figure 5.3: The signal and background distribution of: (g) Azimuthal angular separation between the soft jets and missing transverse energy, (h) Azimuthal angular separation between the $Z\gamma$ system and missing transverse energy, (i) missing transverse energy, (j) Azimuthal angle of the missing transverse energy, (e) missing significant energy and (f) the number of jets.

may be incomplete, inconsistent, or contain many errors. These are some of the challenges that should be overcome before starting modeling. A few techniques such as vectorization, normalization and handling of missing values, and removal of outliers can be used to eliminate these challenges. Vectorization converts categorical values into an integer representation by implementing methods such as one-hot encoding. This makes

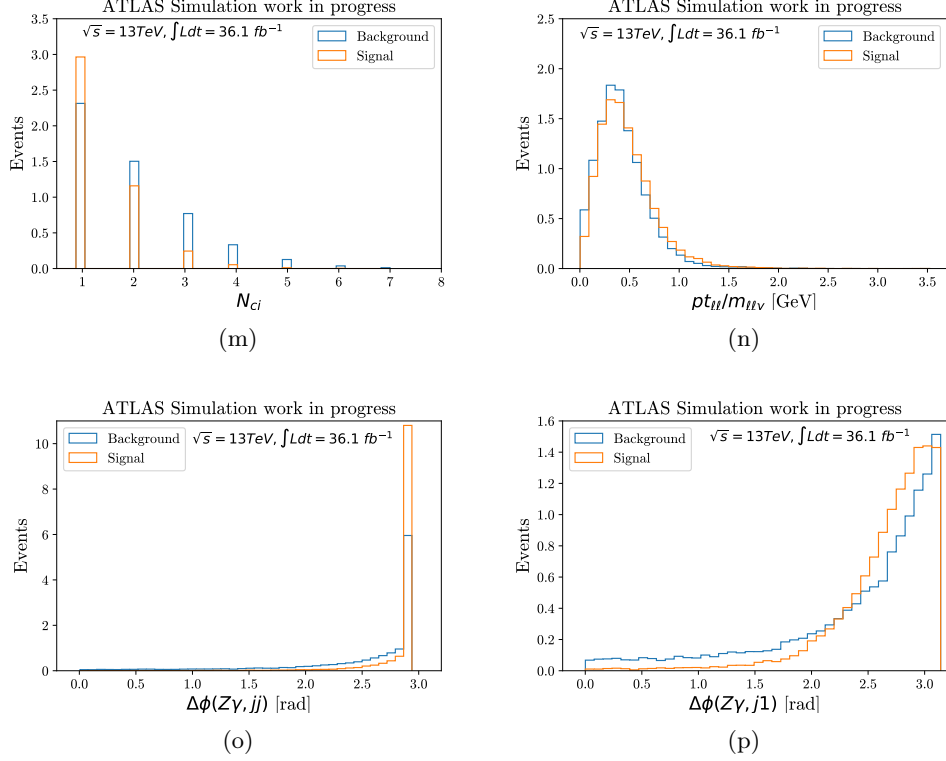


Figure 5.3: The signal and background distribution of: (m) number of central jets, (n) relative pt of a photon, (o) Azimuthal angular separation between the $Z\gamma$ system and the dijet system and (p) Azimuthal angular separation between the $Z\gamma$ system and the leading jet.

it easier for the ML model to handle such features since the model can only input features in numeric format.

Unlike vectorization, normalization is about transforming the data to a similar scale, which helps improve the training data and the performance of the model. If the features are not within a similar scale, for example, one feature is in the range 0 - 1 while the other is within 1000 - 20000. This is considered heterogeneous data and causes the model to have large gradient updates, which in turn prevents the model from converging. This problem can be solved by transforming the data into homogeneous data, scaling all features within the same range without distorting the difference in ranges [216]. As a result, this ensures that the features have the same degree of influence. To achieve this, techniques such as the MinMax scaler can be used, which ensures that the data is within $[0, 1]$. This scaler is defined by:

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}, \quad (5.2)$$

where x_i is the i^{th} entry/record for the variable x , x'_i is the rescaled entry, while $\min(x)$ and $\max(x)$ are the minimum and maximum entries, respectively. This approach was

Parameter	Range	Description
n_estimators	[10 - 100]	Number of trees in the forest
max_depth	[2 - 6]	Maximum tree depth
max_features	[2 - 6]	Maximum number of features to achieve the best split
min_samples_leaf	[2 - 6]	Minimum number of samples needed at the a leaf node
min_samples_split	[2 -5]	Minimum number of samples needed for internal node split
Criterion	[gini, entropy]	Function used to measure the quality of a split

Table 5.2: The parameters used to configure the Random Forest in Python-based package scikit-learn [217].



Figure 5.4: schematic representation of cross-validation.

used to normalize the data for this study. The data is also split into 60:40 for training and testing purposes.

5.4.2 Random Forest Configuration

The RFs for this study are configured using a Python-based package called scikit-learn [218] and optimized using a grid-search algorithm to select the best-performing combination of the parameters listed in Table 5.2. During training, a cross-validation of $k = 5$ is performed, where the dataset is divided into k groups. In this approach, the hold-out technique is implemented for each of the k groups, using one of the groups for testing and the rest for training ($k - 1$). The process is repeated until each group has been used for testing. Figure 5.4 shows the illustration of this technique. This ensures that all data is used for training.

5.4.3 The Deep Neural Networks Configuration

For the DNN configurations, a python API, Keras library with Tensorflow backend [219] was used. Different topological DNN structures were configured for the categories examined in this study. This is to ensure that optimal performance is achieved for each category. Table 5.4.3 shows the range of different hyper-parameters used for the DNN models. The input layer of each DNN consists of 17 neurons representing the kinematic

Hyper-parameter	Range
Input units	17
Hidden layers	[4 - 10]
Hidden nodes	[20 - 250]
Learning rate	$[7.72 \times 10^{-6} - 7.52 \times 10^{-2}]$
weight decay	$[1.03 \times 10^{-8} - 1.03 \times 10^{-1}]$
Dropout	[0 - 0.5]
Epochs	[50 - 500]
Activation function	[ReLU, sigmoid]

Table 5.3: The hyper-parameters used to configure different DNN models.

features of the dataset. The choice of the activation function implemented has a significant impact on the training dynamics and performance of the model [197]. Therefore, the activation functions should be computationally efficient as they are computed hundreds or thousands of times, especially in the case of DNNs.

ReLU has been the most widely used activation function in DL [197, 220–222] since its introduction by Nair and Hinton in 2010. One of the advantages of this function over other activation functions such as sigmoid is that it rectifies the vanishing problem and is less computationally intensive since it does not require exponential and division calculations [223]. It is mostly used for the hidden layers in DNNs. However, for the output layer, the sigmoid function is preferred in DNNs because of its saturation [224]. Ref. [197] indicates that DNNs are easier to optimize with ReLU activation function than with other activation functions such as sigmoid or tanh. Moreover, the ReLU activation function given by equation 5.3 simplifies the training process of DNNs by mitigating the challenges of weight-initialization and vanishing gradient [225]:

$$f(x) = \begin{cases} 0 & \text{for } x \leq 0, \\ x & \text{for } x > 0. \end{cases} \quad (5.3)$$

For this reason, ReLU was chosen for the hidden layers of DNNs in this study, while sigmoid was used for the output layer. Even though ReLU is considered the best and most commonly used activation function. One of its drawbacks is that it easily overfits, *i.e.*, the model fits perfectly to the training data and fails dismally when generalized to a new data set. It is important to ensure that the performance of the model measured on the training set is close to the performance on the validation set. If the performance on the training set is significantly higher, then the model is overfitting. This happens when the algorithm becomes overly sensitive or captures noise in the data, resulting in low bias and high variance. A more complex model on a small amount of data can also lead to this problem. However, if the model is too simple, it fails to capture the underlying trend in the training data [92]. This leads to high bias and low variance. For a perfect model, there must be a balance, because both overfitting and underfitting will result in poor prediction of the test data. The following steps are some of the approaches considered in this study to mitigate this problem. It is worth noting that hyperparameter

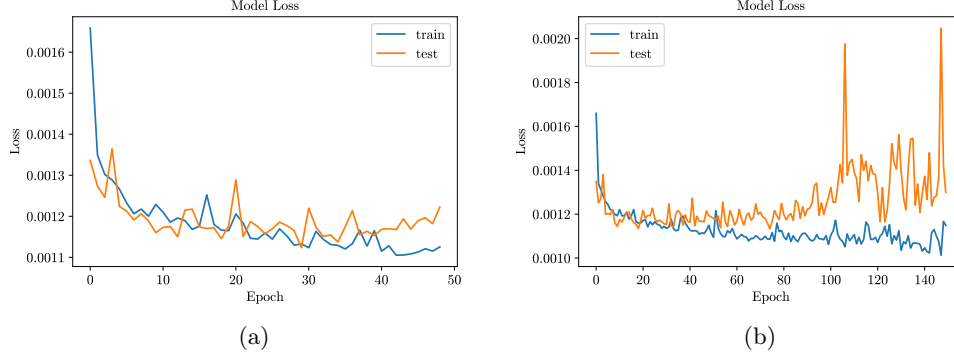


Figure 5.5: Model loss as a function of epochs for: (a) DNN with an implementation of the early stopping parameter and (b) DNN without the early stopping parameter .

optimization was performed to ensure optimal performance for each configured model.

1. **Early stopping:** Machine learning models tend to overfit when trained for longer iterations/epochs than required. This can be illustrated by plotting the training error or loss graph against the number of epochs, which indicates the error based on the training and validation sets. In this case, an early stop is imposed as soon as the error in the validation set stops decreasing and starts increasing. In this way, the optimal model is saved while the performance of the model is measured at each iteration. The model starts overfitting immediately after this step. Ref. [226] points out that early stopping is the simplest and most commonly used way to stop overfitting, and therefore helps in selecting models that provide better generalization [227, 228]. This is illustrated in Figure 5.5(a), where (b) shows training until all epochs are exhausted. From Figure 5.5(b), it is easy to see that the model starts overfitting after only 50 epochs. This parameter not only provides optimal results but is also computationally efficient since the optimal generalization performance is achieved with fewer training iterations. This is illustrated based on the VBF category for the mass point of 200 GeV.

2. Dropout:

Dropout mitigates the overfitting problem by reducing interdependent learning between neurons [229] and preventing them from co-adapting too much [226]. This technique is called dropout because it randomly drops neurons from specific layers along with their connections during training. During this dropout phase, the outputs of the dropped neurons in a given run, i.e. forward or backward pass, cannot be considered. Refs. [229, 230] indicate that this technique is a simple and efficient way to prevent overfitting and that it also allows for better performance of DNN models compared to other regularization methods. In DNNs, this is represented by a hyperparameter p , which indicates the probability that a hidden node of the network structure is preserved.

3. Learning Rate

This is one of the hyperparameters of DNNs, which is used to update the weights

of DNNs with respect to the loss gradient. Consequently, this parameter determines the rate at which DNN updates the already learned concept. Similar to the complexity of the DNN models, which affects performance, the choice of learning rate is also important. A smaller learning rate takes time to converge, thus longer a training time [231, 232]. This is due to the smaller adjustments to the weights at each update. However, a higher learning rate leads to rapid changes, resulting in a sub-optimal set of weights. This is arguably the most important hyperparameter when configuring a DNN architecture [233]. It is therefore important to choose an efficient number for this parameter, which of course is mostly on a trial-and-error basis. This is given by a small positive number that is in the range between 0 and 1.

The conventional and reasonable approach is that the learning rate should decrease monotonically during training [234] because tuning and shaping of this parameter are important for training performance [235]. An additional parameter is introduced to gradually reduce the learning rate during training. This parameter is called weight decay. It is used to update the learning rate using the following equation:

$$\alpha = \alpha_0 \times \frac{1}{1 + \text{weight decay} \times \text{epoch}}, \quad (5.4)$$

where α is the learning rate for the current epoch, α_0 is the initial learning rate, decay is a positive number, and epoch is the current update number. This is calculated after each mini-batch. The learning rate and weight decay hyperparameters implemented in this study are listed in Tables 5.4.3.

5.5 Performance

Training a DNN and RF produces classifiers capable of separating signal and background events. The classifiers are trained with events labeled with the ground truth, *i.e.*, 0 for background and 1 for signal, in the case of fully supervised learning. In the case of weakly supervised learning, the same approach is followed, but 0 is assigned to the background in the sideband region, while 1 is assigned to the mixture of signal and background in the mass window region. During the training phase, the classifier assigns a value for the event, a value that represents the probability that the event is either signal or background. Overall, this helps in evaluating how well the classifier is performing.

Figures 5.6 and 5.7 show the DNN and RF response distribution functions, respectively, for signal and background for two categories, namely $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ and $N_{jets} \geq 2$; $\Delta\eta_{jj} > 2$; $m_{jj} > 300 \text{ GeV}$. These results are the outcome of implementing the proposed methods described in Section 4.2.3 and 4.2.4. Both approaches appear to provide significant separation performance, with fully supervised learning showing much better performance. A Significant separation between signal and background is seen with weakly supervised learning, where the model was trained on the mixture of signal and background, but evaluated with pure signal and background (see Figure 5.6). However, this does not apply to weakly supervised learning on RF, which shows minimal separation between the signal and the background.

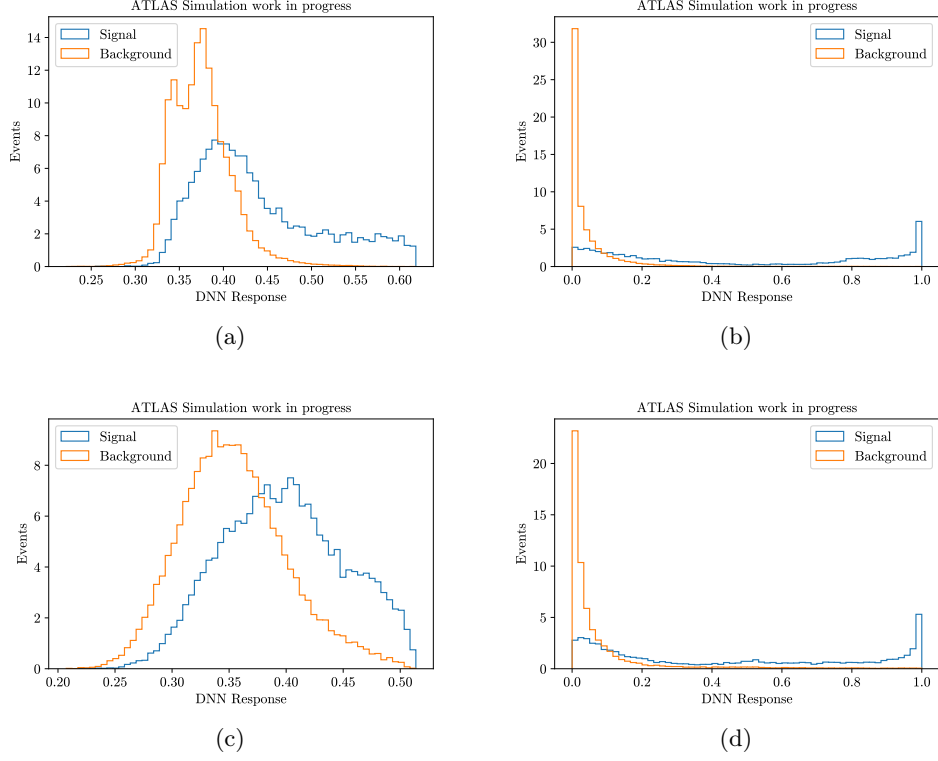


Figure 5.6: The DNN response distributions: (a) Weakly supervised learning for $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category, (b) Fully supervised learning for $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category, (c) Weakly supervised learning for $N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$ category and (d) Fully supervised learning for $N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$ category.

In general, fully supervised learning gives much better performance for all categories, and the remaining categories are shown in the Appendices C and B for DNNs and RFs, respectively. The fully supervised learning shows excellent discrimination between signal and background, with the distributions showing a long tail at both ends, especially for background, implying that the probability of rejecting background events is high. Even though the weakly supervised learning gives significant separation for DNNs, there still appears to be an overlap between signal and background.

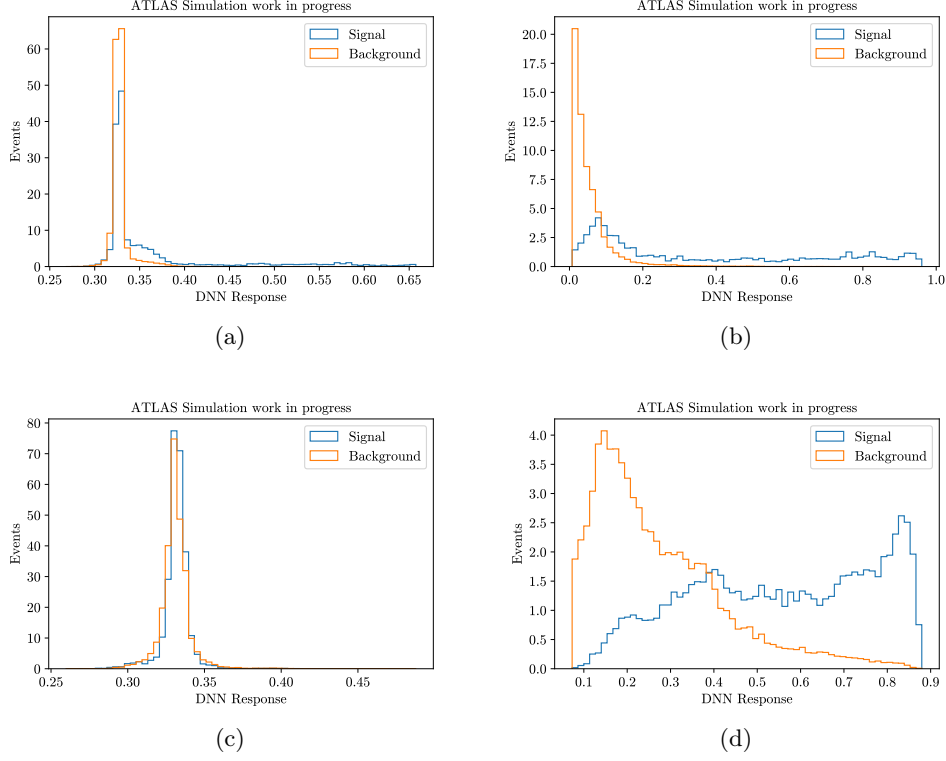


Figure 5.7: The RF response distributions: (a) Weakly supervised learning for $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category, (b) Fully supervised learning for $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category, (c) Weakly supervised learning for $N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$ category and (d) Fully supervised learning for $N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$ category.

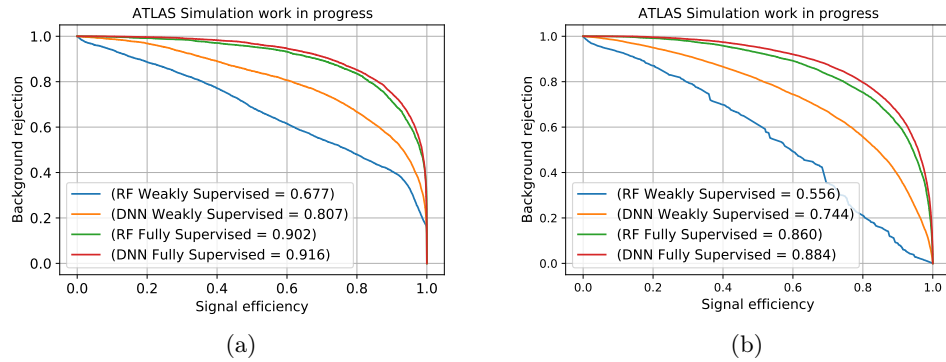


Figure 5.8: The ROC curves for: (a) $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$, (b) $N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$.

The results of the trained DNNs and RFs are evaluated using a ROC curve given by the signal efficiency versus background rejections. This is achieved by using AUC as the

Mass Point (GeV)	Process	Category	Random Forest (%)	Deep Neural Network (%)
150	Wh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	64.6	75.0
	Wh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	51.4	66.4
	Zh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	69.4	82.1
	Zh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	47.1	66.7
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	61.4	74.6
200	ggF	Inclusive	53.2	54.7
	Wh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	67.7	80.7
	Wh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	51.7	62.0
	Zh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	65.1	84.7
	Zh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	53.4	64.1
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	55.6	74.4
300	ggF	Inclusive	52.8	57.1
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	49.0	77.3
500	ggF	Inclusive	47.6	59.6
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	63.1	79.6

Table 5.4: A comparison of the performance of DNNs and RFs in the form of ROC AUC based on weakly supervised learning.

Mass Point (GeV)	Process	Category	Random Forest (%)	Deen Neural Networks(%)
150	Wh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	88.4	90.6
	Wh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	83.6	87.3
	Zh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	89.7	91.3
	Zh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	83.7	87.5
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	85.8	88.0
200	ggF	Inclusive	62.6	68.7
	Wh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	90.2	91.6
	Wh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	84.5	88.6
	Zh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	91.5	92.6
	Zh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	84.5	87.9
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	86.0	88.4
300	ggF	Inclusive	65.4	71.5
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	87.7	88.7
500	ggF	Inclusive	71.7	72.5
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	90.4	91.4

Table 5.5: A comparison of the performance of DNNs and RFs in the form of ROC AUC based on fully supervised learning.

primary metric which is given by the integral of the ROC curve. The larger the AUC value, the better the accuracy of the performance. The ROC curves shown in Figure 5.8 give the quantified performance of the DNN and RF classifiers for both the weakly supervised learning and weakly supervised learning approaches. This is given by the signal efficiency versus background rejection.

The summary of all these AUCs can be found in Tables 5.4 and 5.5 for weakly supervised learning and fully supervised learning, respectively. From these summaries, it can be seen that weakly supervised learning is not efficient when applied to the RF in comparison to DNN. However, the performance of the two algorithms is quite comparable

for the fully supervised learning approach with RF being slightly lower than the DNN in some cases. This is in agreement with the results in Ref. [236]. It is for this reason that the entire study is carried out to completion with the implementation of DNN. For reference, the ROC curves corresponding to these results are provided in Appendix E. The two techniques are used for different processes at the respective mass points. The performance of the DNN for the 200 GeV mass point for inclusive is significantly worse than that of the other categories. This suggests that the optimized DNN was not successful in separating the signal from the background.

5.5.1 Significance

The ability of the classifier can be read directly from the distribution curves, where it can be seen how the classifier managed to separate the background from the signal. From this DNN response distribution is derived the significance, the standard metric in HEP. This significance is calculated based on the background rejection using the following equation:

$$\text{significance} = \frac{S}{\sqrt{S+B}}, \quad (5.5)$$

where S is the number of signal events and B is the number of background events. The significance indicates the maximum signal-to-noise ratio achieved by the DNN classifier. Figures in Appendix D shows the significance results for all categories and their respective mass points. These results are summarized in Tables 5.6 and 5.7 for fully supervised learning and weakly supervised learning, respectively.

Maximum significance appears to be achieved at 0% background rejection rate for the $N_{jets} \geq 2$; $60 < m_{jj} < 120$ GeV category for 200 GeV mass point. The same effect is observed for the inclusive category for 200, 300, and 500 GeV. This is due to the fact that the DNN classifier does not achieve a clearer separation between the signal and the background. The results of this significance can be seen in the plots presented in Appendix D. The maximum significance along with the percent rejection of the background is shown in Tables 5.6 and 5.7 for all categories.

Mass Point (GeV)	Process	Category	Maximum significance	Background rejection (%)
150	Wh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	2.72	95
	Zh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	4.55	99
	Wh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	1.97	10
	Zh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	1.97	20
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	2.25	70
200	ggF	Inclusive	1.89	0
	Wh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	2.94	99
	Zh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	4.25	99
	Wh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	1.88	0
	Zh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	1.89	10
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	2.14	70
300	ggF	Inclusive	1.96	0
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	2.22	80
500	ggF	Inclusive	1.90	0
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	2.01	70

Table 5.7: Maximum significance for various background rejections from the DNN response distribution with weakly supervised learning.

Mass Point (GeV)	Process	Category	Maximum significance	Background rejection (%)
150	Wh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	4.74	99
	Zh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	5.33	99
	Wh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	3.49	95
	Zh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	3.16	95
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	3.45	95
200	ggF	Inclusive	2.09	40
	Wh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	4.16	99
	Zh	$S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$	4.65	99
	Wh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	3.28	95
	Zh	$N_{jets} \geq 2; 60 < m_{jj} < 120 \text{ GeV}$	2.91	90
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	3.14	95
300	ggF	Inclusive	2.29	60
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	2.45	90
500	ggF	Inclusive	2.13	40
	VBF	$N_{jets} \geq 2; \Delta\eta_{jj} > 2; m_{jj} > 300 \text{ GeV}$	2.61	90

Table 5.6: Maximum significance for various background rejections from the DNN response distribution with fully supervised learning.

5.6 Spurious Studies

In this section, we estimate the uncertainty in the choice of different functions to describe the background shape. This uncertainty relates to the fitting of a pure background template using a given signal and background function with a non-zero N_{sig} result. This is called a spurious signal, intended to select the optimal functional form [18], where N_{sig} denotes the number of signal events. The properties of the fit of each tested function are presented in Table 5.8 for the VBF category. Table 5.8 summarises the results shown in Figures 5.9 - 5.11. The rest of the other categories are shown in Appendix F. The background template is a combination of background MC samples and the inverse ID data sample ($Z + jets$) in the following categories: VBF, $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$;

Function	S/ Δ S in %	Spurious signal	N par	Chi2 Prob	Pass selection
Category-0					
FK0	72.3	18.6	2	14.8	PASS
FK1	136	38.7	3	11.9	PASS
Category-1					
FK0	67.7	16.5	2	24.7	PASS
FK1	125	31.8	3	21.6	PASS
Category-2					
FK0	73	17.1	2	23.1	PASS
FK1	102	26.1	3	21	PASS
Category-3					
FK0	74.5	16	2	33.6	PASS
FK1	108	26.8	3	33.1	PASS
Category-4					
FK0	78.9	15	2	23.8	PASS
FK1	124	26.3	3	21.7	PASS
Category-5					
FK0	60.2	11.2	2	17.4	PASS
FK1	94.3	18	3	16.1	PASS
Category-6					
FK0	59.9	-8.32	2	13.3	PASS
FK1	79.8	12.9	3	11.7	PASS

Table 5.8: Fitted spurious signal yield at 150 GeV for VBF phase-space. The spurious signal yield is normalized to 140 fb^{-1} . Category-0 represents no background rejection, whereas Category-1 to Category-6 represent 10%, 20%, 30%, 40% and 50% background rejection.

$N_{jets} \geq 2; 75 \text{ GeV} \geq m_{jj} \geq 105 \text{ GeV}$ and $N_{jets} \geq 1$. There are the different orders of the so-called dijet functions (FKO):

$$f_k(x; b, \{a_k\}) = (1 - x)^b x^{\sum_{j=0}^k a_j \log(x)^j}, k = 0, 1, 2, \quad (5.6)$$

denoted as FK0, FK1 and FK2. This is done to test various functions for each functional form, with different fits tested by varying the position of the signal peak. The template sample is stored in the histogram with 1 GeV per bin from 130 to 1000 GeV. A scan of the presence of the signal is performed with a step of 1 GeV, and the waveform varies as a function of mass. To decide which functional form describes the background form well, an "F-test" procedure is introduced, using a higher-order functional form if its "F-test" gives a p-value of less than 5%.

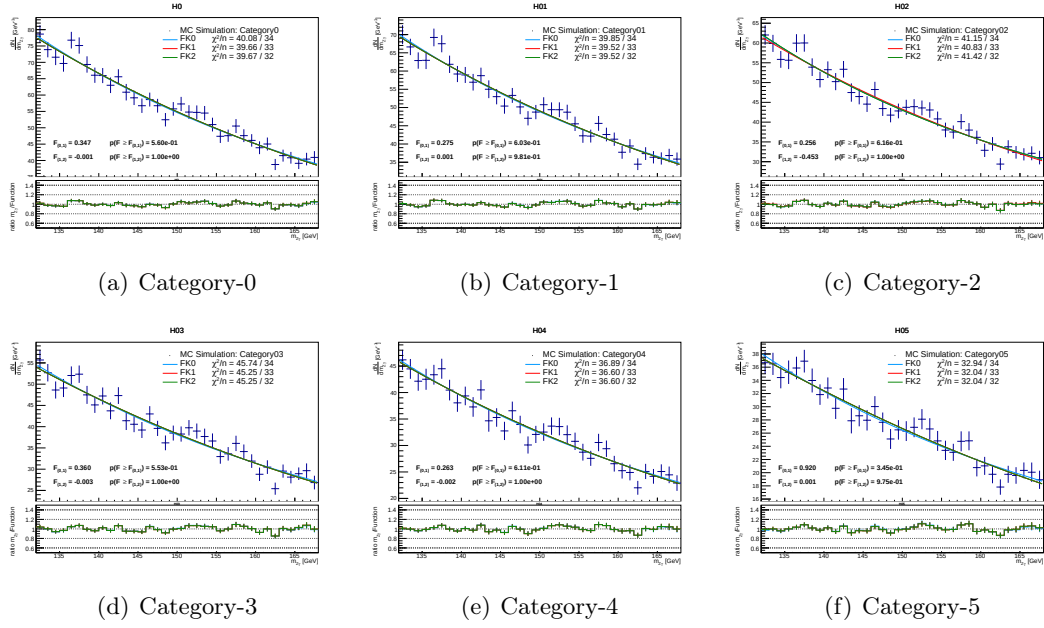


Figure 5.9: Fitting on the background MC template VBF 150 GeV.

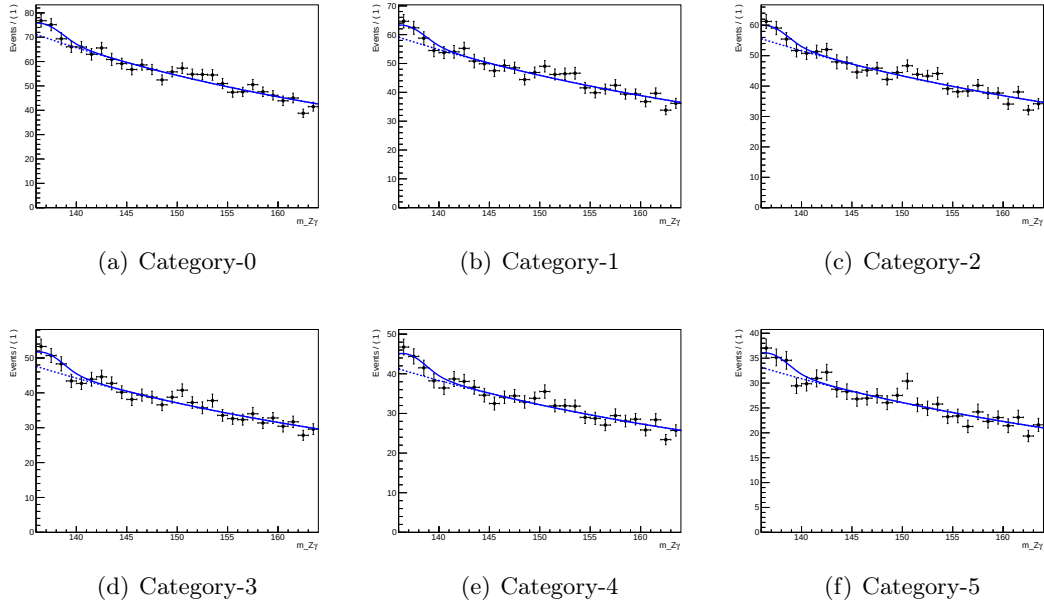


Figure 5.10: Fitting on the background template for VBF 150 GeV, with the S+B function.

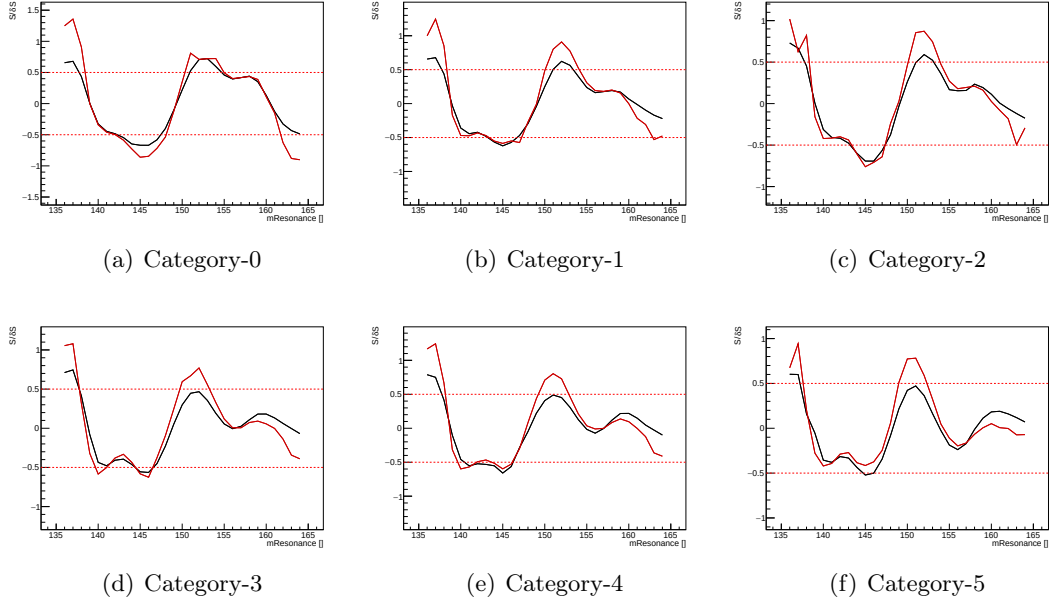


Figure 5.11: The $N_{spurious}/\delta N_{spurious}$ as a function of mass for each functional form for VBF 150 GeV.

5.7 Signal Parameterization

The signal parameterization is performed on all the mass points for the categories considered for this study. This is done where the signal mass distribution is modelled from MC simulations as the sum of a double-sided crystal ball (DSCB) function [89], described as follows.

$$N \cdot \begin{cases} e^{-t^2/2}, & \text{if } -\alpha_{Lo} \leq t \leq \alpha_{Hi}, \\ \frac{e^{0.5\alpha_{Lo}^2}}{[\frac{\alpha_{Lo}}{n_{Lo}}(\frac{\alpha_{Lo}}{n_{Lo}} - \alpha_{Lo} - t)]^{n_{Lo}}}, & \text{if } t < -\alpha_{Lo}, \\ \frac{e^{0.5\alpha_{Hi}^2}}{[\frac{\alpha_{Hi}}{n_{Hi}}(\frac{\alpha_{Hi}}{n_{Hi}} - \alpha_{Hi} + t)]^{n_{Hi}}}, & \text{if } t > \alpha_{Hi}, \end{cases} \quad (5.7)$$

where $t = \Delta m_x / \sigma_{CB}$, $\Delta m_x = m_x - \mu_{CB}$, N is the normalization parameter, μ_{CB} is the peak of the Gaussian distribution, σ_{CB} represents the width of Gaussian part of the function, while $\alpha_{Lo}(\alpha_{Hi})$ is the point where the Gaussian becomes a power-law on low (high) mass side and $n_{Lo}(n_{Hi})$ is the exponent of the power-law. The μ_{CB} and σ_{CB} for the VBF category are summarized in Table 5.9 and shown in Figure 5.13 as a function of the invariant mass. The rest of the other categories are presented in Appendix G. While the results of the DSCB fits for inclusive are shown in Fig. 5.12. The data invariant mass shape is provided by the following extended likelihood formula:

$$\mathcal{L}_c = \text{Pois}(n_c | N_c(\theta)) \cdot \prod_{i=1}^{n_c} f_c(m_{\ell\ell\gamma}^i, \theta) \cdot G(\theta), \quad (5.8)$$

Mass Point (GeV)	Background rejection (%)	μ_{CB} (GeV)	σ_{CB} (GeV)
150	10	149.71 ± 0.06	2.30 ± 0.08
	20	149.70 ± 0.04	2.32 ± 0.04
	30	149.71 ± 0.03	2.35 ± 0.04
	40	149.74 ± 0.05	2.28 ± 0.09
	50	149.71 ± 0.04	2.30 ± 0.05
200	inc	199.71 ± 0.02	3.02 ± 0.02
	10	199.68 ± 0.03	2.93 ± 0.04
	20	199.67 ± 0.03	2.92 ± 0.05
	30	199.67 ± 0.04	2.91 ± 0.05
	40	199.66 ± 0.04	2.92 ± 0.04
	50	199.65 ± 0.03	2.97 ± 0.03
300	inc	299.41 ± 0.03	4.23 ± 0.02
	10	299.56 ± 0.06	3.87 ± 0.08
	20	299.54 ± 0.06	3.88 ± 0.08
	30	299.55 ± 0.06	3.86 ± 0.08
	40	299.57 ± 0.06	3.87 ± 0.08
	50	299.59 ± 0.06	3.85 ± 0.09
500	inc	499.48 ± 0.05	5.06 ± 0.09
	10	499.34 ± 0.08	5.19 ± 0.15
	20	499.36 ± 0.09	5.20 ± 0.15
	30	499.37 ± 0.09	5.27 ± 0.15
	40	499.45 ± 0.10	5.22 ± 0.17
	50	499.42 ± 0.09	5.27 ± 0.21

Table 5.9: Summary of μ_{CB} and σ_{CB} values for VBF signal samples at $N_{jets} \geq 2$, $m_{jj} > 300$ GeV and $\Delta\eta_{jj} > 2$, after background rejection.

where G denotes the terms constraining the nuisance parameters, f and c denote the $m_{\ell\ell\gamma}$ invariant mass shape and different categories, respectively. While n_c represents the observed number of events in a specific category c , and $N_c(\theta)$ is the expected number of events in category c determined by the model parameters θ .

The first term in the likelihood function is the Poisson probability $\text{Pois}(n_c | N_c(\theta))$, which quantifies the probability of observing n_c events given an expected number of events $N_c(\theta)$. This term accounts for the statistical fluctuations in the data.

The second term in the likelihood function is a product over all events in the category c , denoted as $\prod^{n_c} f_c(m_{\ell\ell\gamma}, \theta)$. This term represents the probability density function (PDF) f_c for each event's invariant mass $m_{\ell\ell\gamma}$ in category c . The PDF captures the shape of the invariant mass distribution and is parameterized by the model parameters

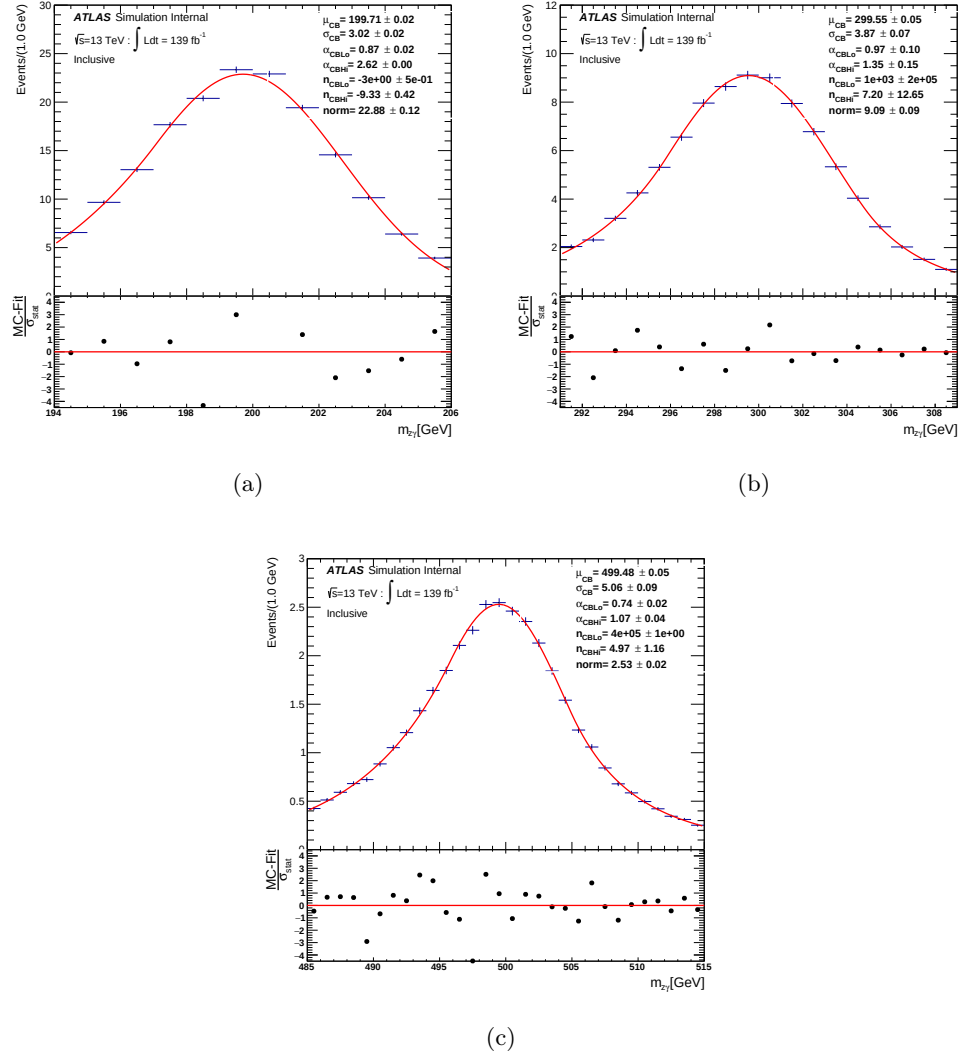


Figure 5.12: Results of individual fit of DSCB parameters as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for VBF inclusive: (a) 200 GeV (b) 300 GeV and (c) 500 GeV.

θ . The product is taken over all observed events in category c .

The third term, denoted as $G(\theta)$, incorporates the effects of nuisance parameters in the analysis. Nuisance parameters are additional parameters that describe uncertainties or systematic effects in the measurement. This term includes constraints on the nuisance parameters to ensure that they are consistent with prior knowledge or auxiliary measurements. These constraints are typically represented as Gaussian or log-normal probability distributions.

By maximizing the likelihood function $\mathcal{L}_c$ with respect to the model parameters θ , we can determine the parameter values that best describe the observed data in terms of

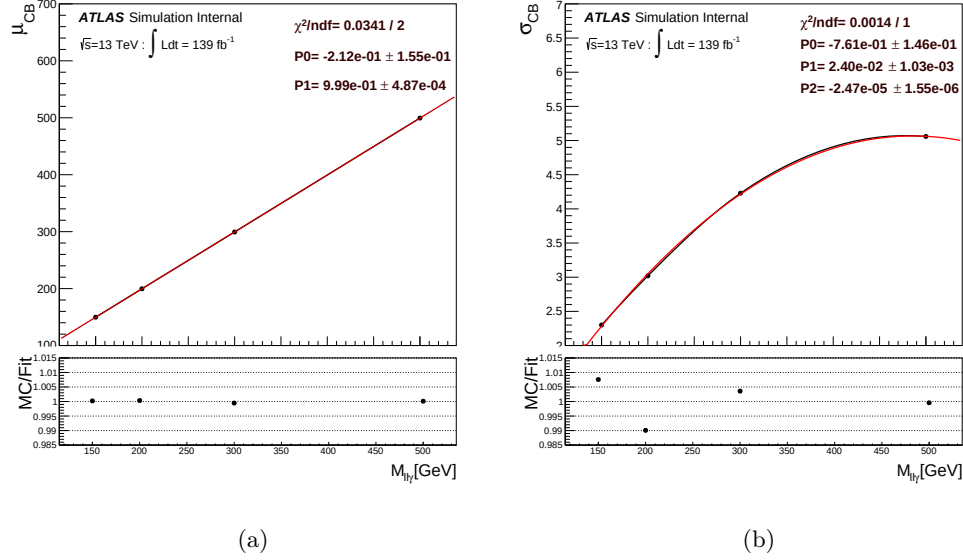


Figure 5.13: Parameterization of: (a) μ_{CB} and (b) σ_{CB} as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for VBF

the invariant mass spectrum. This maximization procedure takes into account both the statistical fluctuations in the data and the shape of the invariant mass distribution .

The extended likelihood formula enables a comprehensive analysis of the invariant mass spectrum, considering both the event yields and their shape. By combining the statistical and shape information, this approach provides a robust and accurate characterization of the observed data, facilitating the study of resonant peaks and their properties in this research.

5.8 Discussion and Conclusions

In this analysis, the discriminative ability of the weakly supervised learning technique is evaluated and compared to that of fully supervised learning. The ATLAS Monte Carlo samples are used to evaluate the two techniques. This corresponds to the non-resonant $Z\gamma$ dataset, which is the dominant background since it accounts for more than 90% of the total background. The signal represents the simulated Higgs-like to $Z\gamma$ final state [237].

The performance of the DNN is compared to that of the RF, both of which are based on weakly and fully supervised learning. The DNN gives better results in each case of the learning techniques, with fully supervised learning RF being comparable to the DNN. However, this is not the case for weakly supervised learning, where a significant difference was found between the two algorithms. The results seem to be different for the different categories considered, with the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ giving optimal results for different mass points than the other categories. The difference between these categories

can be attributed to their specific event configurations.

It is worth noting that different machine learning algorithms perform differently for different applications, which is due to the specific applications and the dimensionality of the dataset. Both algorithms are capable of modeling both linear and nonlinear complex functions, but the results seem to indicate that the DNNs have higher potential. The DNNs are selected for further analysis in this study based on their performance. The combination of DNNs with weakly supervised learning and fully supervised learning is crucial for HEP because they are able to show improvements in classification that can provide further evidence for some of the BSM theories. This suggests that DNNs are useful in extracting the discriminative information in the physical objects for the purpose of classification.

This work focuses on finding resonances using ML techniques such as weakly supervised learning and fully supervised learning in conjunction with DNNs. The envisioned methodology has been tested on the production of the Higgs boson in the $\gamma\gamma$ study discussed in Chapter 4, and furthermore has been used in the $Z\gamma$ final state with the intention of setting up a search for new phenomena in high mass final states for the LHC. The results show that weakly supervised learning is capable of extracting signals from the background, but its performance is somewhat lower than that of fully supervised learning. DNNs have shown their strength in learning from highly complex data and discovering nonlinear feature combinations. The advancement of deep learning techniques could help in exploring the high volumes of data in HEP with tiny amounts of signal that could be hiding under large amounts of backgrounds.

Chapter 6

Conclusions

Advanced machine learning techniques have emerged as a necessity for complex data analysis in HEP. Such techniques, especially Deep Learning, have proven to be useful in event reconstruction and identification. In this work, Deep Learning in the form of DNNs based on fully supervised and weakly supervised learning is used to search for heavy resonances at the EW scale with topological requirements. The SM Higgs boson with di-photon decay is used to evaluate the performance of the proposed method. This is presented in Chapter 4 with the sole purpose of evaluating the capability of this methodology, which is then used to search in the $Z\gamma$ final states in Chapter 5. The study helped to understand how well ML algorithms perform when dealing with particles produced with different production mechanisms. Thus, supervised learning with an unlabeled signal achieved better performance than fully supervised learning for a signal sample when the different production methods were left unlabeled. Weakly supervised learning with an unlabeled signal sample is also compared with these two methods. The weakly supervised learning with unlabeled signal somewhat achieves a significantly worse performance compared to the other approaches.

The evaluation of the $Z\gamma$ final state is performed using the Monte Carlo simulated signal samples for 139 fb^{-1} integrated luminosity for Run 2 at the LHC. The study is performed in both inclusive and exclusive regions of phase space tailored to a particular production mode. It was conducted in the 150 - 600 GeV range, with the overall goal of establishing a search for the high mass points at the LHC. The results of this study suggest that weakly supervised learning is not as efficient as fully supervised learning in the search. Thus, the performance of fully supervised learning outperforms that of weakly supervised learning. The performance is driven by the event selections or configurations. The results of the SM Higgs boson with di-photon decay are shown in Ref. [1], with those of the $Z\gamma$ final state in [2].

The signals and topologies used in this study can also be used to search for heavier or lighter SM Higgs-like bosons. However, this is only true if the topological constraints are loose enough. Then they can be used for more general searches for bosons. Other signatures, such as those of the MSSM (minimal supersymmetric model), can be added. For example, in the presence of a b -tagged jet, it is well motivated to use weakly supervised

learning, requiring a maximum number of jets to remove potential signals associated with top quarks. Other types of searches, such as searches for resonances in multiparticle cascades, can also be performed using this method.

In contrast to the generic search, the study of phase space is guided by physical considerations. Consequently, the concept of signal optimization and predicted importance, which is deeply rooted in search with fully supervised learning, must be modified and adapted to this technique. The model coverage or the ability of the search to cover large parts of the parameter space should be emphasized. In this study, weakly supervised learning is defined in the context of using sidebands to represent the background. This method can also be used to control large samples as well as signatures that do not have bounded resonances.

Appendices

Appendix A

Correlation Matrix

Correlation matrices for the production mechanisms taken into consideration for this study are provided in Appendix A as supplementary material. For instance, Figures A.1 and A.2 depict, in the inclusive category for 200 GeV, respectively, the correlation between pairs of features for background and signal. For this category, it appears that some of the pairs of features, such as when m_{jj} paired with features like η_{j2} , N_j , N_{cj} and $\Delta\phi(Z\gamma, jj)$, show a strong correlation than others. Additionally, the individual correlation between ϕ_{ET}^{miss} , $pt_{\ell\ell}/m_{\ell\ell\gamma}$, and $\Delta\phi(softjets, E_T^{miss})$ and the remaining features is insignificant. Figures A.3 and A.4 for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category for background and signal in 200 GeV, respectively, show a nearly identical trend. The $N_{jets} \geq 2$; $60 < m_{jj} < 120$ GeV category (Figure A.5 - A.6), in comparison, appears to have insignificant correlations between most of the features, in addition to a few with a strong correlation such as (η_{j1}, η_{j2}) , $(E_T^{miss}, S_{ET}^{miss})$, and $(\Delta\phi(Z\gamma, j1), \Delta\phi(Z\gamma, jj))$, due to their degree of similarity.

A.1 Inclusive

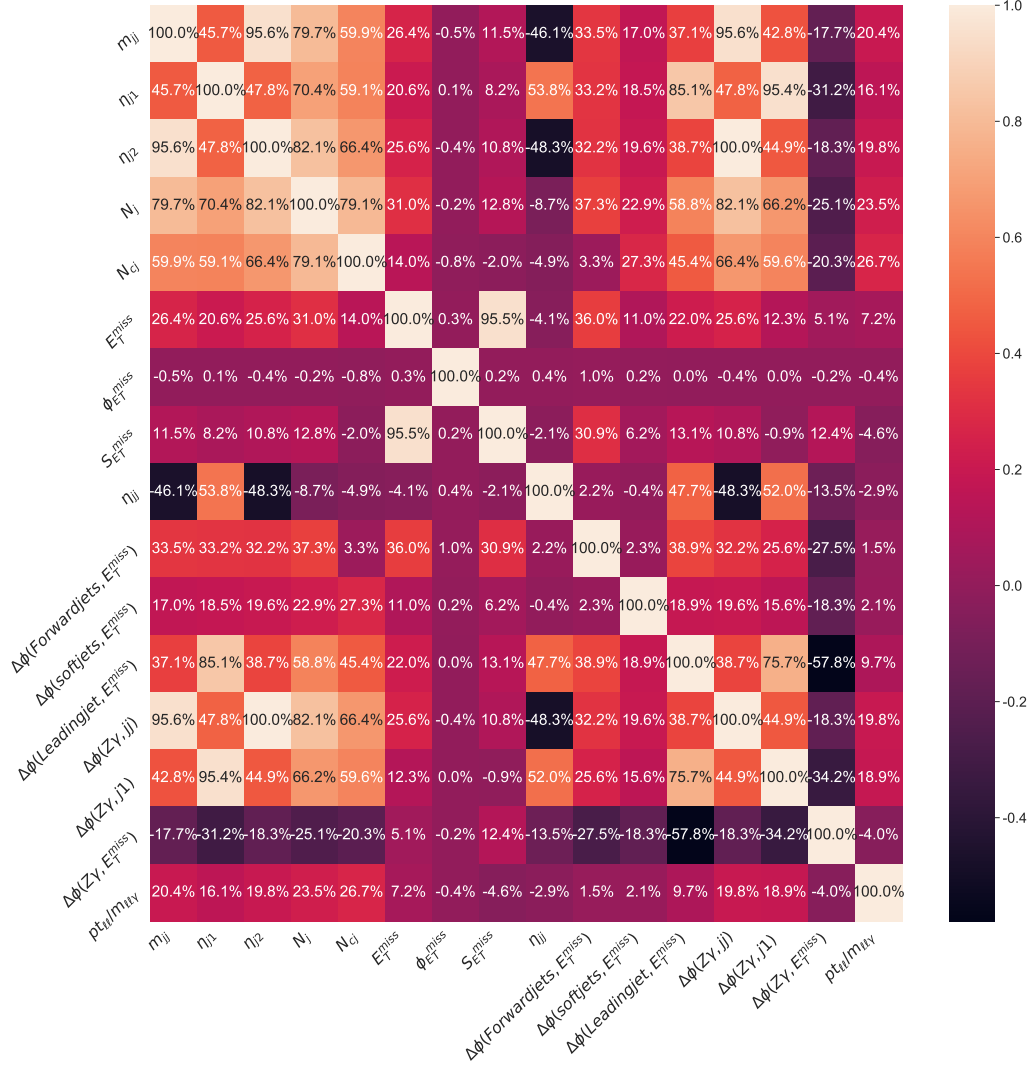


Figure A.1: Correlation matrix of the implemented features/kinematics for the 200 GeV background in the inclusive category.

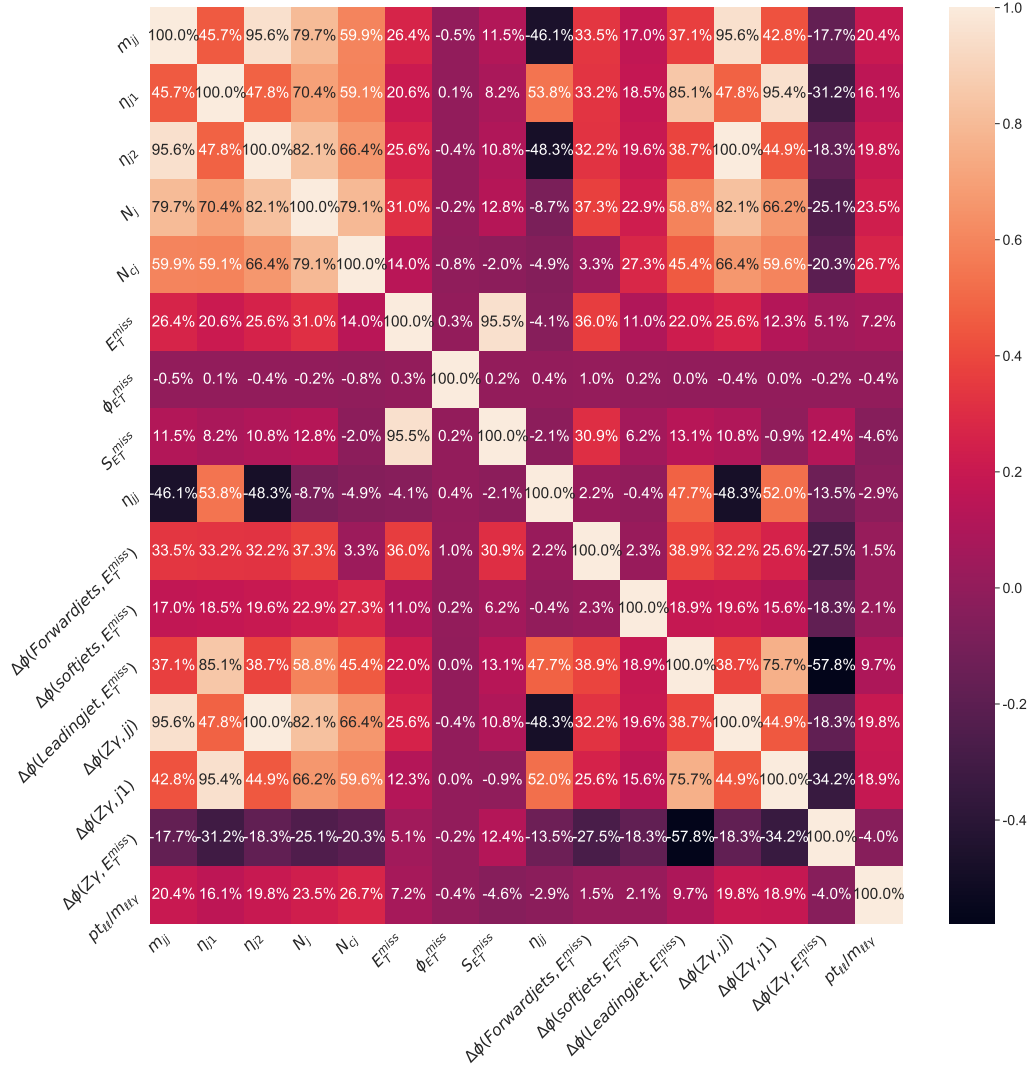


Figure A.2: Correlation matrix of the implemented features/kinematics for the ggF signal for 200 GeV in the inclusive category.

A.2 $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$

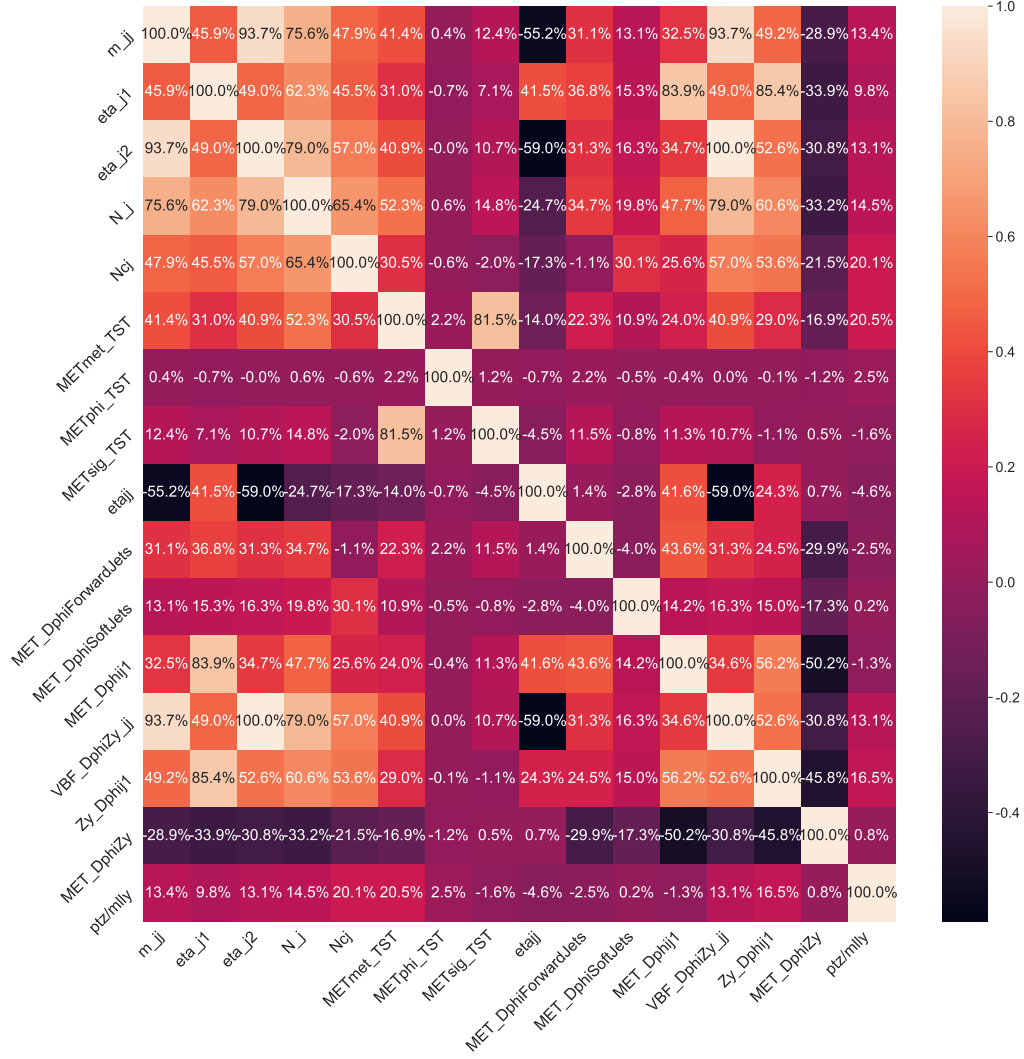


Figure A.3: Correlation matrix of the implemented features/kinematics for the 200 GeV background in the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category.

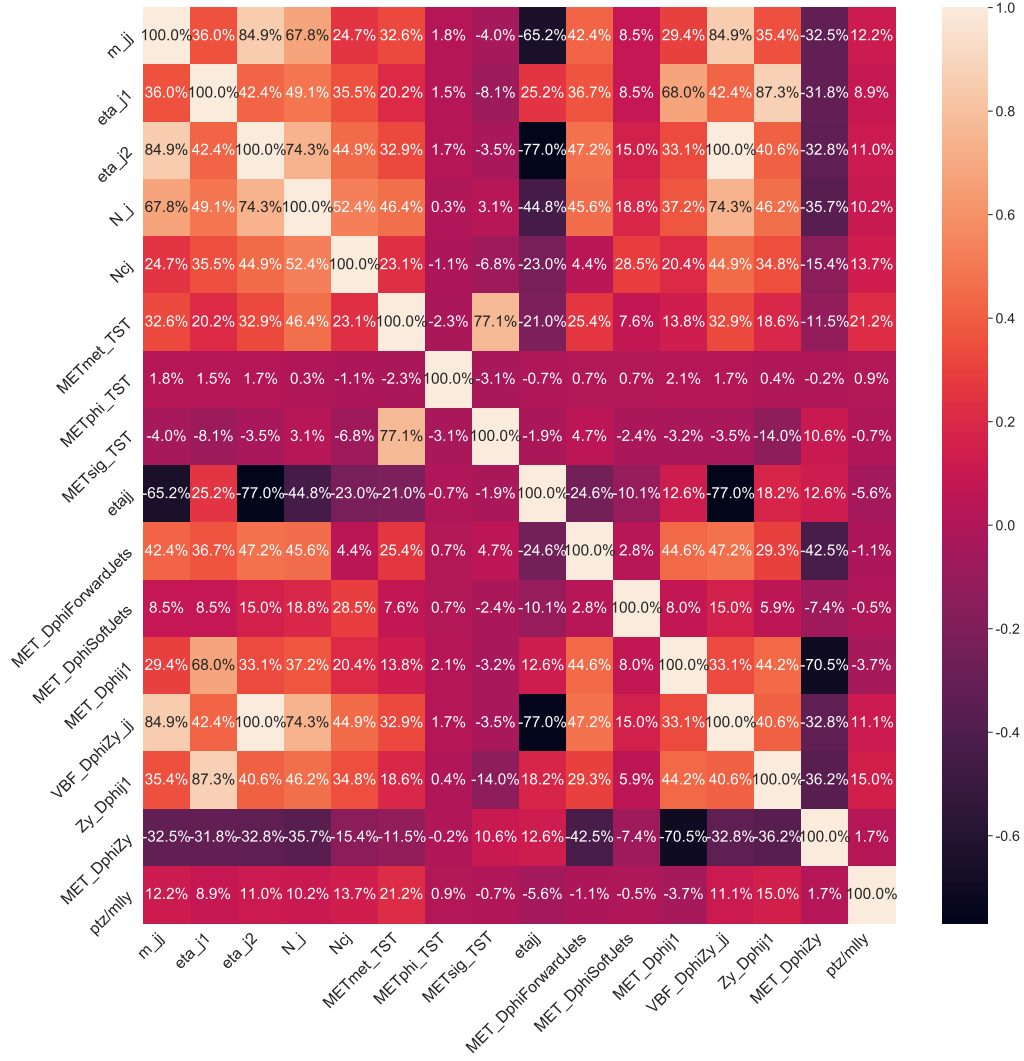


Figure A.4: Correlation matrix of the implemented features/kinematics for the Wh signal for 200 GeV in the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category.

A.3 $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV

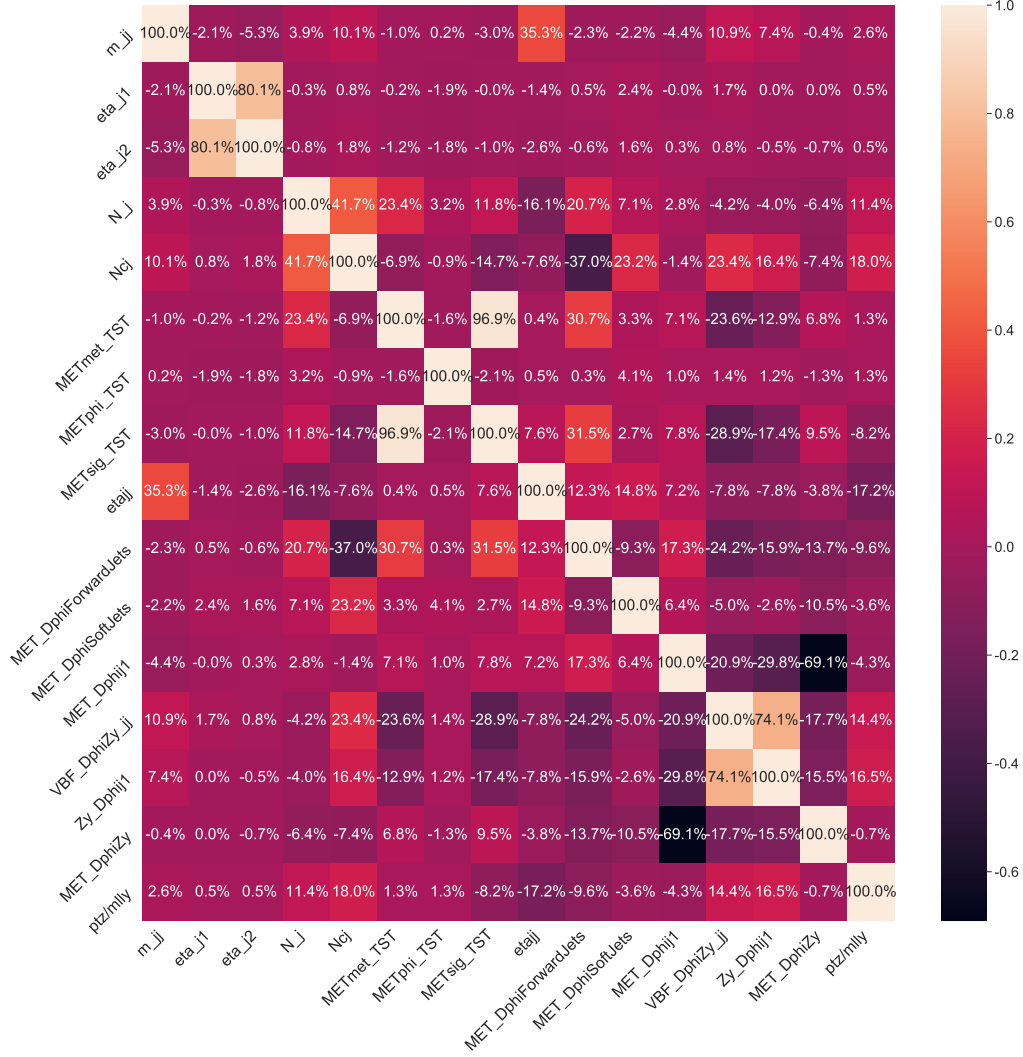


Figure A.5: Correlation matrix of the implemented features/kinematics for the 200 GeV background in the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV category.

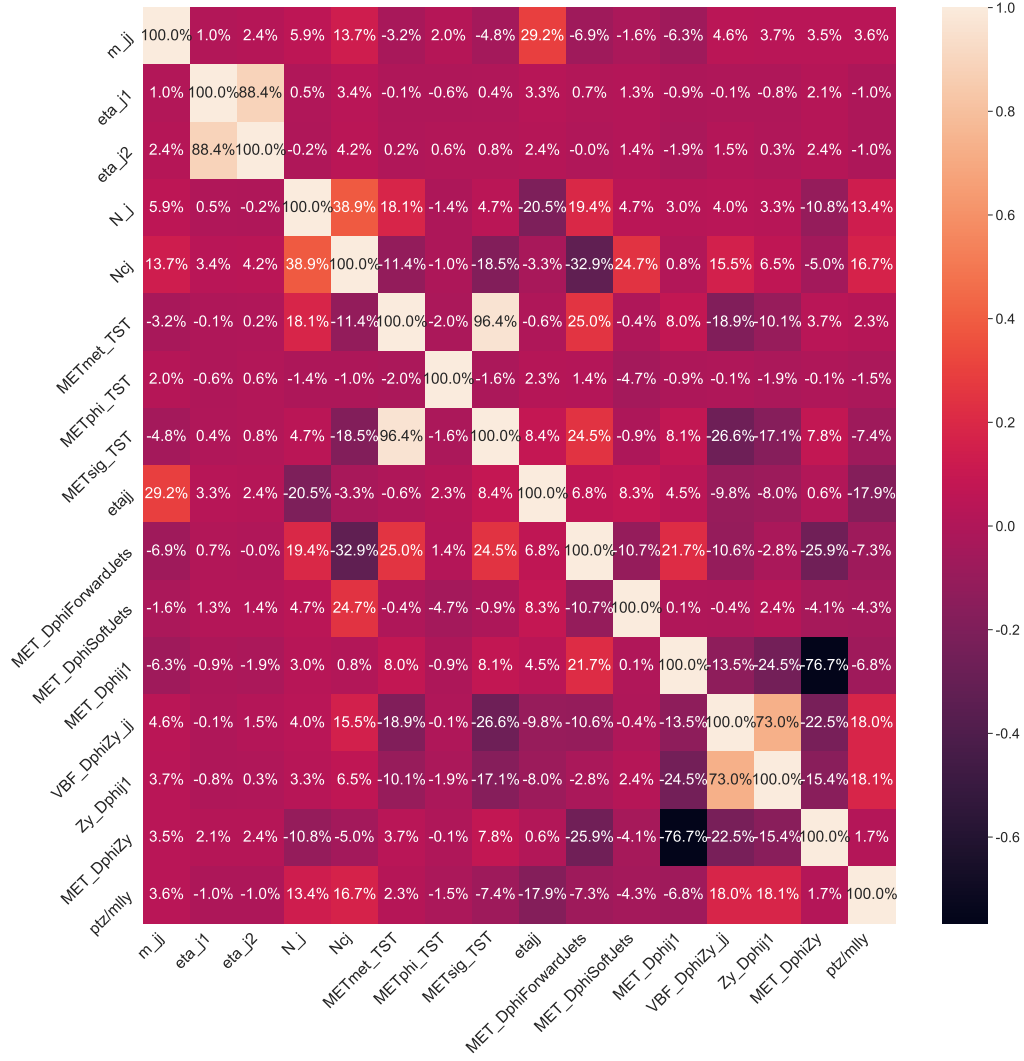


Figure A.6: Correlation matrix of the implemented features/kinematics for the Zh signal for 200 GeV in the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV category.

Appendix B

Random Forest Response Distributions

Appendix B includes the Random Forest results for both weakly supervised learning and fully supervised learning discussed in Section 5.5. These results are grouped into categories and their respective mass points. For example, the results for the inclusive category for mass points 200, 300, and 500 GeV are shown in Figures B.1 - B.3. Compared to fully supervised learning, the weakly supervised learning approach appears to have achieved a minimal separation between the background and signal. In particular, for the 500 GeV mass point, there appears to be significant separation between signal and background for fully supervised learning. This is also evident in the other categories, from Figure B.4-B.11. In the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category, the weakly supervised learning shows slightly better performance than the inclusive case.

B.1 Inclusive

B.1.1 200 GeV

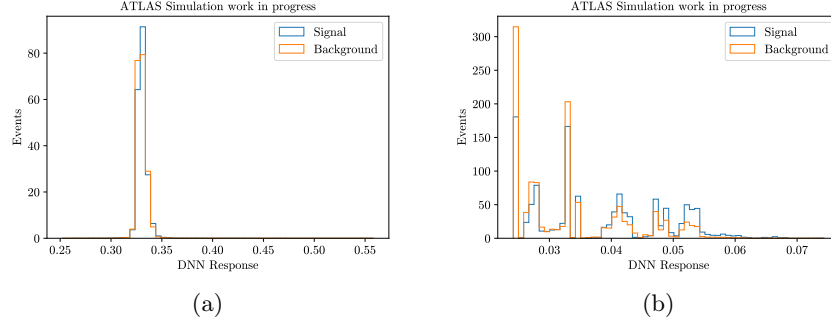


Figure B.1: RF Response Distribution for the inclusive category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

B.1.2 300 GeV

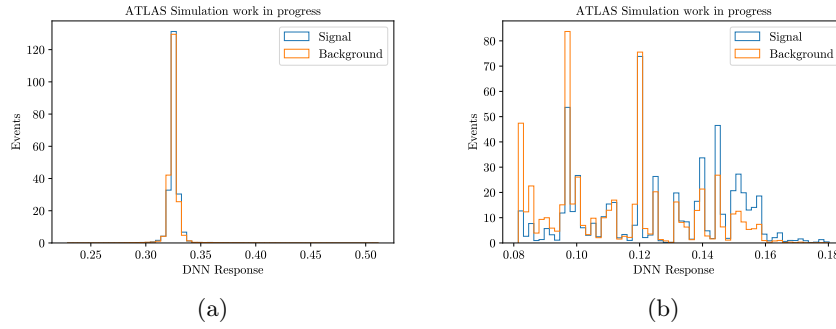


Figure B.2: RF Response Distribution for the inclusive category for 300 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

B.1.3 500 GeV

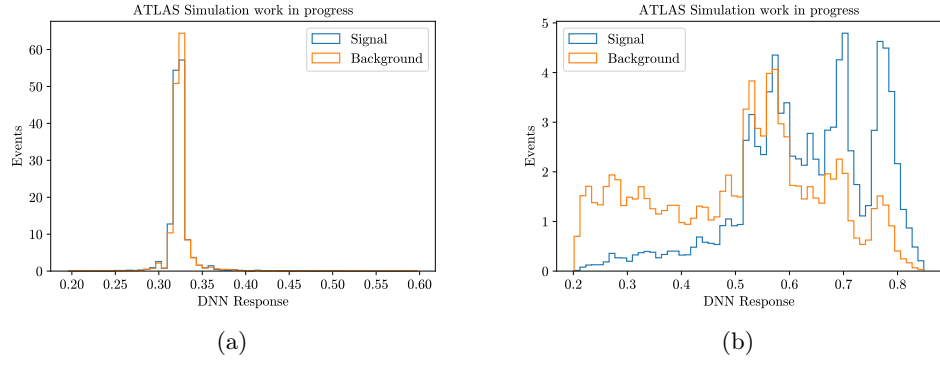


Figure B.3: RF Response Distribution for the inclusive category for 500 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

B.2 $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$

B.2.1 150 GeV

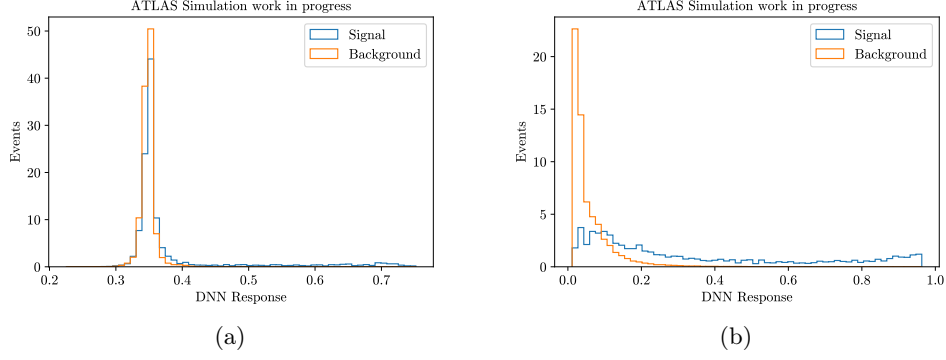


Figure B.4: RF Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

B.2.2 200 GeV

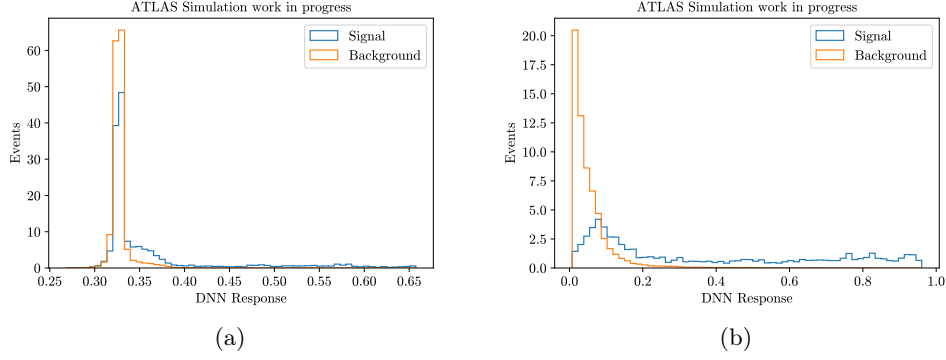


Figure B.5: RF Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$ category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

B.3 $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$

B.3.1 150 GeV

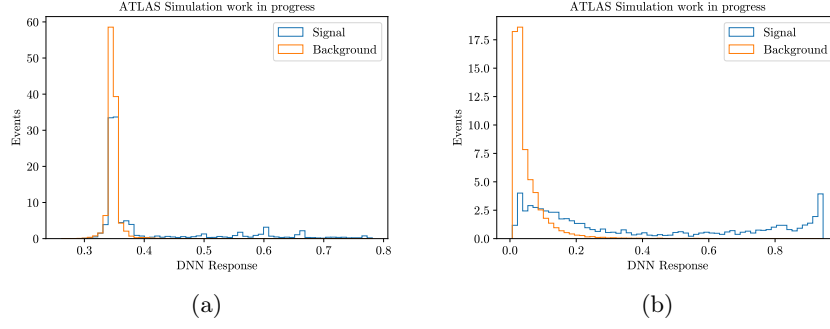


Figure B.6: RF Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

B.3.2 200 GeV

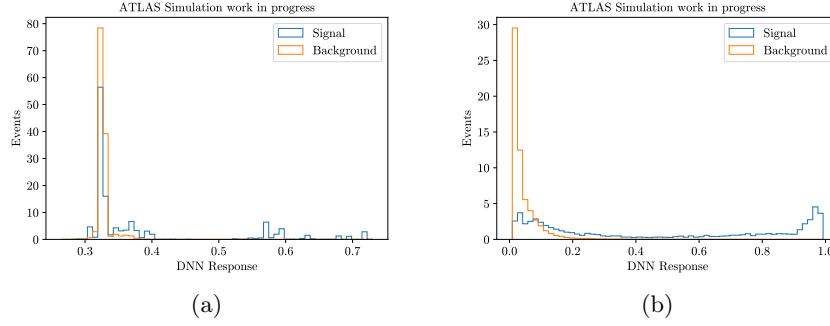


Figure B.7: RF Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

B.4 $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh

B.4.1 150 GeV

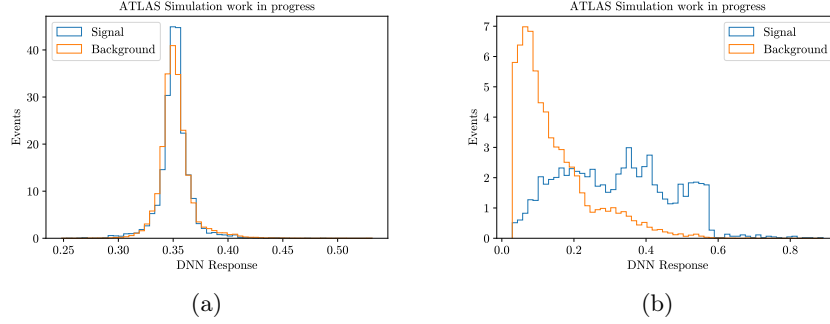


Figure B.8: RF Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

B.4.2 200 GeV

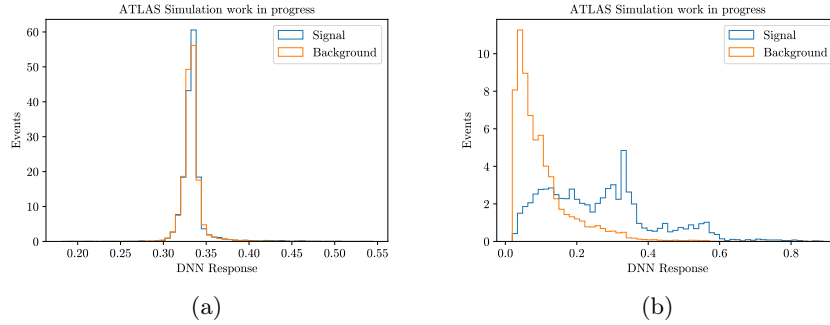


Figure B.9: RF Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

B.5 $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh

B.5.1 150 GeV

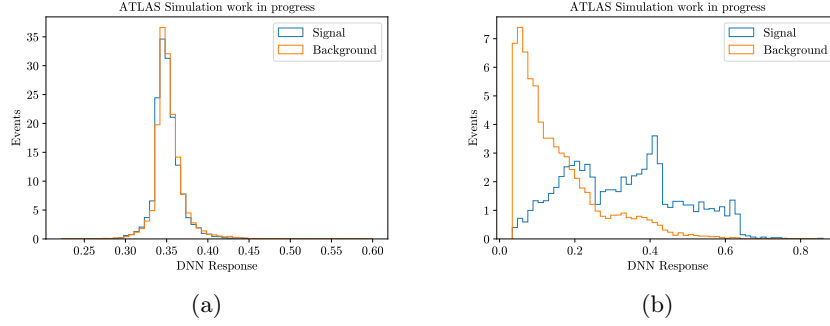


Figure B.10: RF Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

B.5.2 200 GeV

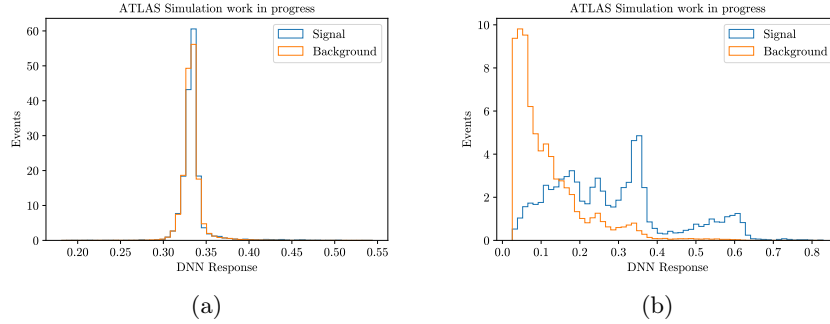


Figure B.11: RF Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

Appendix C

DNN Response Distributions

Appendix C includes the DNN results for both weakly supervised learning and fully supervised learning discussed in Section 5.5. These results are organized into categories and their respective mass points. For instance, Figures C.1 - C.3 show the results for the inclusive category for mass points of 200, 300, and 500 GeV, respectively. Compared to fully supervised learning, the weakly supervised learning approach appears to have only partially separated the background from the signal in each case. In particular, for the 500 GeV mass point, there appears to be significant separation between signal and background. The other categories in Figures C.4 - C.15 also show this. It is important to note that the results of the two techniques are significantly better than those of the counterpart, in the form of RF in Appendix B.

C.1 Inclusive

C.1.1 200 GeV

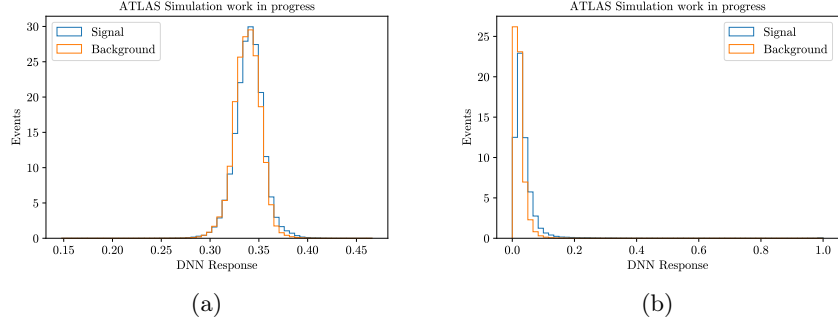


Figure C.1: DNN Response Distribution for the inclusive category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.1.2 300 GeV

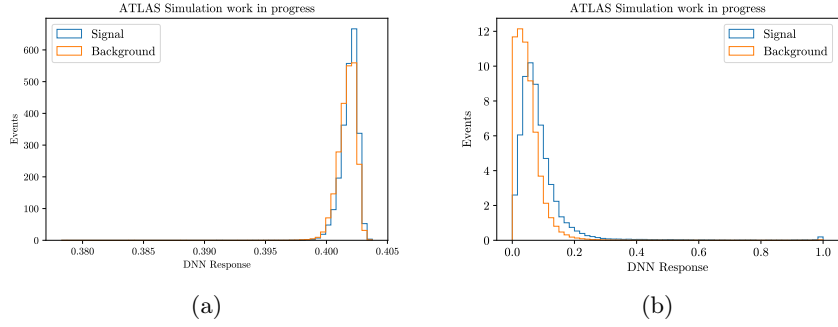


Figure C.2: DNN Response Distribution for the inclusive category for 300 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.1.3 500 GeV

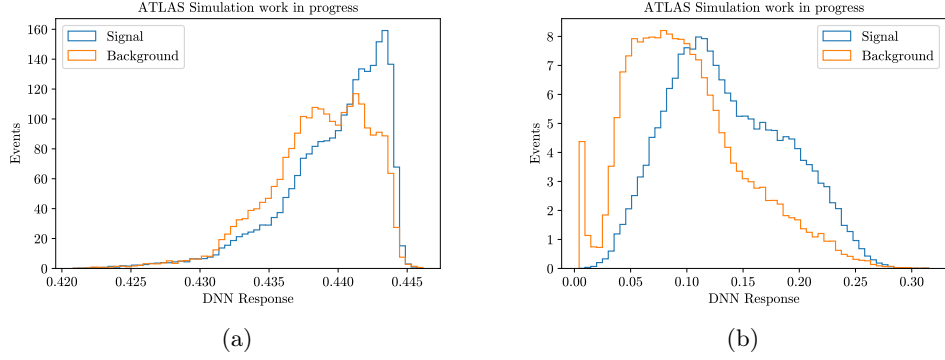


Figure C.3: DNN Response Distribution for the inclusive category for 500 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.2 $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$

C.2.1 150 GeV

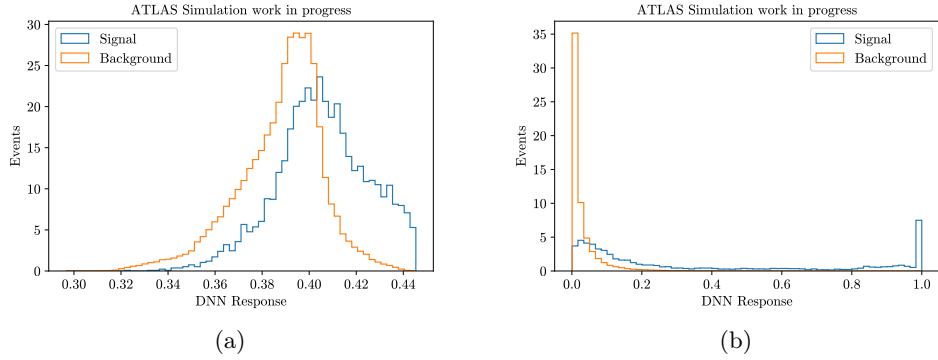


Figure C.4: DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.2.2 200 GeV

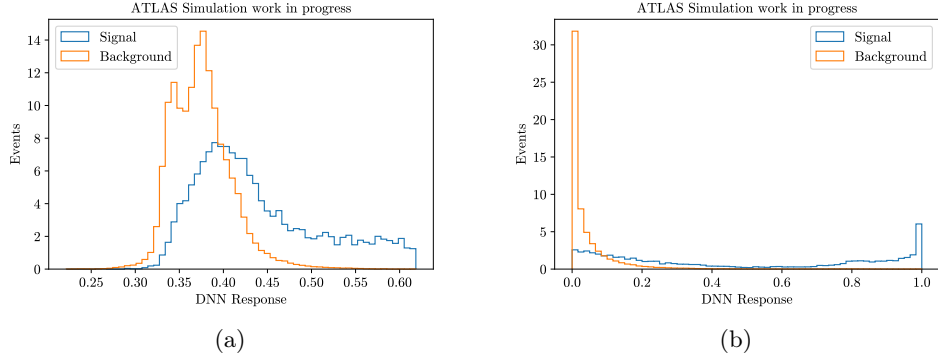


Figure C.5: DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$ category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.3 $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$

C.3.1 150 GeV

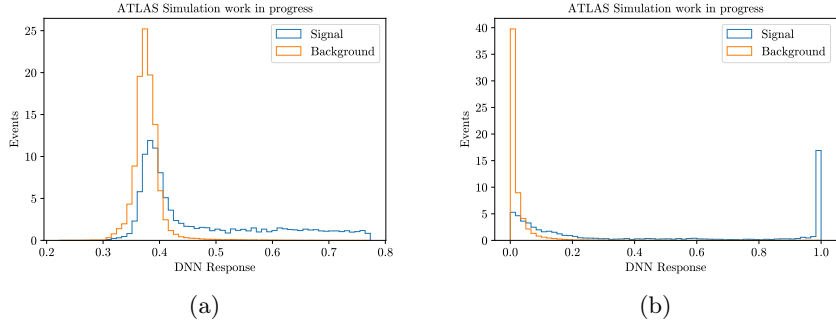


Figure C.6: DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.3.2 200 GeV

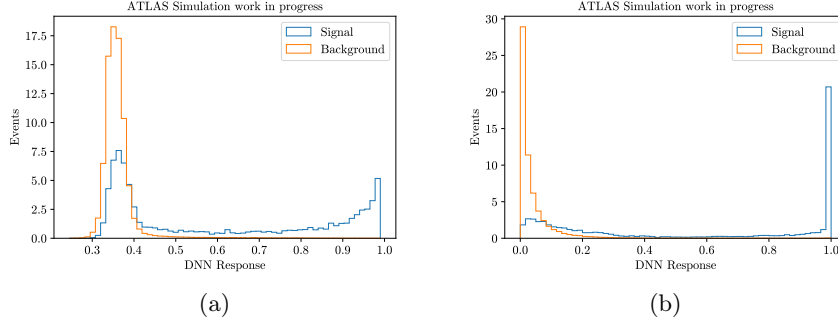


Figure C.7: DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.4 $N_{jets} \geq 2$ and $60 < m_{jj} < 120 \text{ GeV}, Wh$

C.4.1 150 GeV

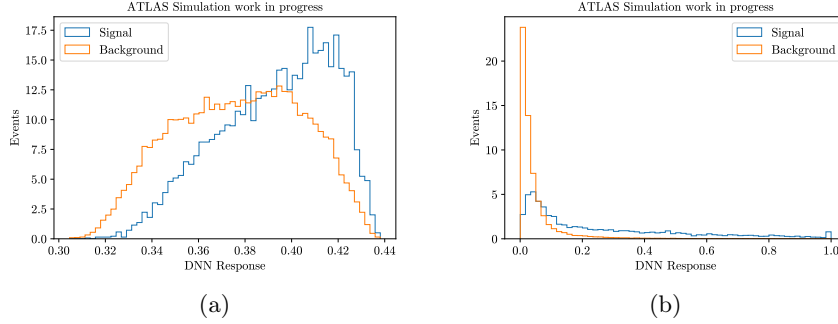


Figure C.8: DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120 \text{ GeV}, Wh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.4.2 200 GeV

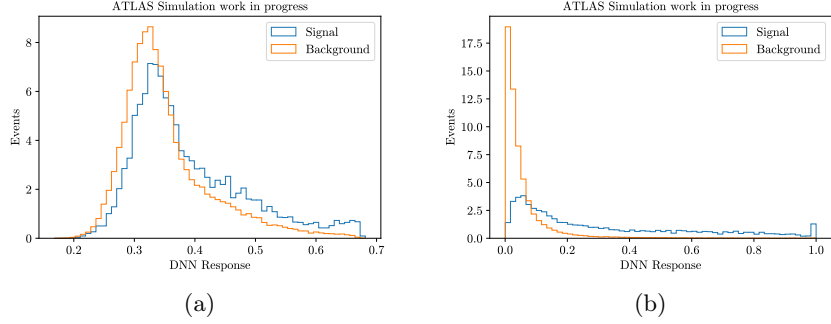


Figure C.9: DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.5 $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh

C.5.1 150 GeV

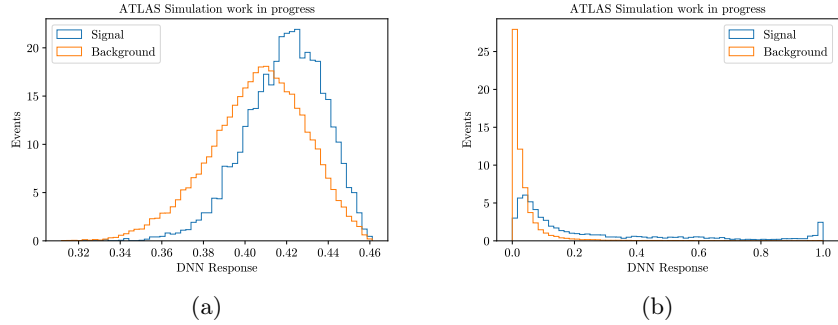


Figure C.10: DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.5.2 200 GeV

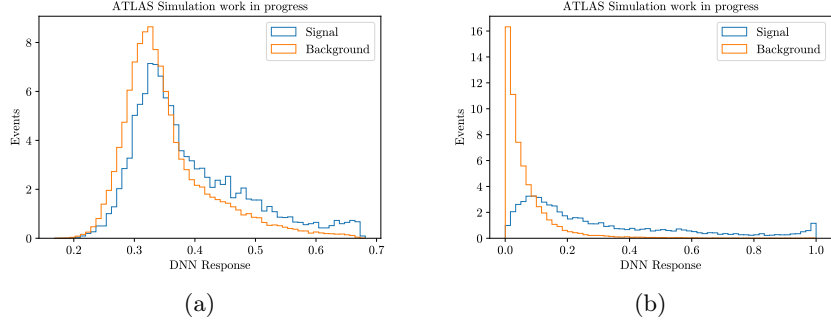


Figure C.11: DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.6 VBF

C.6.1 150 GeV

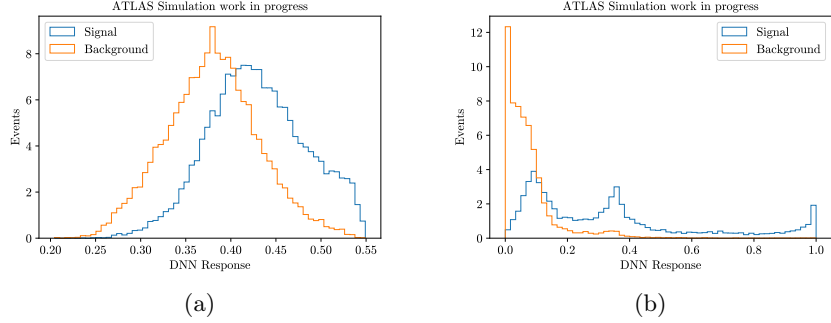


Figure C.12: DNN Response Distribution for the VBF category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.6.2 200 GeV

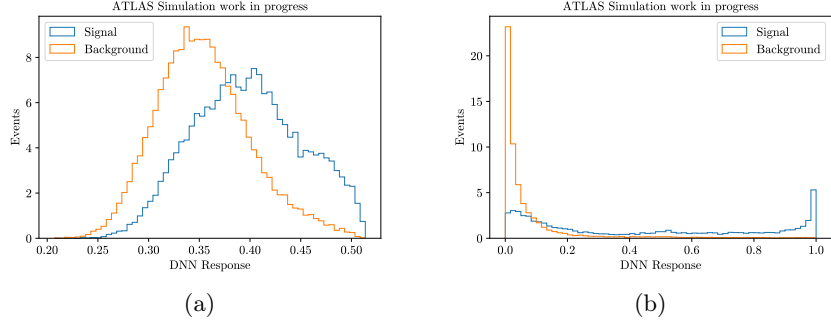


Figure C.13: DNN Response Distribution for the VBF category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.6.3 300 GeV

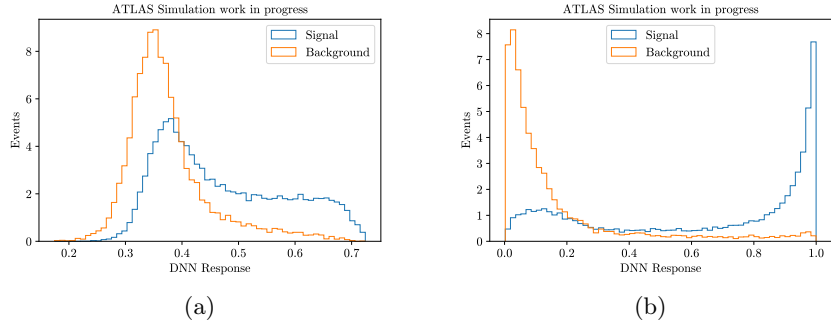


Figure C.14: DNN Response Distribution for the VBF category for 300 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

C.6.4 500 GeV

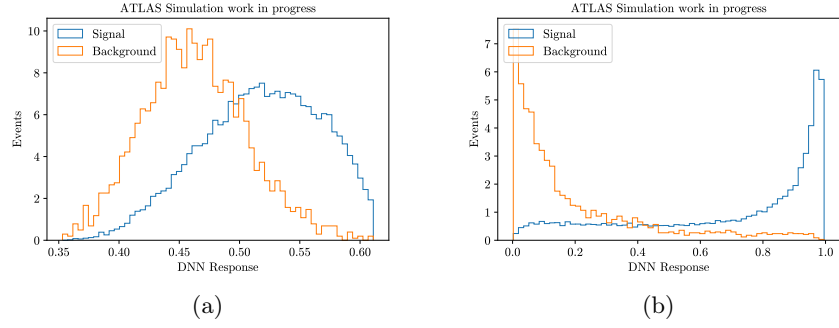


Figure C.15: DNN Response Distribution for the VBF category for 500 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

Appendix D

Significance graphs

Significance was derived based on the distributions shown in Appendix C, which are calculated based on background rejection. These plots were used to quantify the amount of background that must be rejected to achieve the maximum signal-to-noise ratio. Figure D.1(a) shows that maximum significance was achieved at 0% background rejection. This is an indication that the classifier cannot distinguish or separate the signal from the background. This is the same case for all weakly supervised learning approaches in the inclusive category. In the cases where significant separation has been achieved, the maximum significance can be anything but 0% rejection, as shown in the remaining plots, except for the weakly supervised learning for the $N_{jets} \geq 2$; $60 < m_{jj} < 120$ GeV at 150 GeV and the fully supervised learning for 200 GeV.

D.1 Inclusive

D.1.1 200 GeV

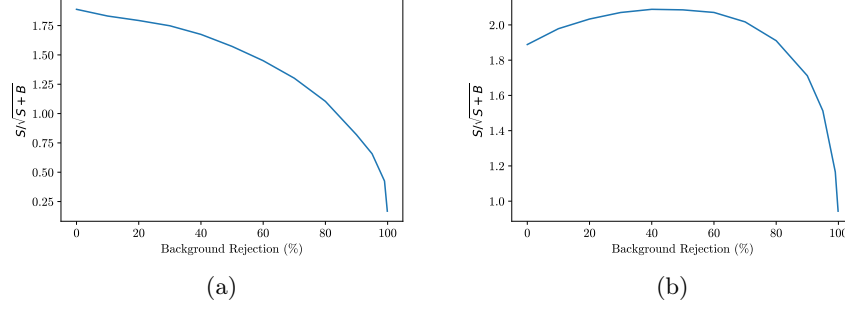


Figure D.1: DNN Response Distribution for the inclusive category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.1.2 300 GeV

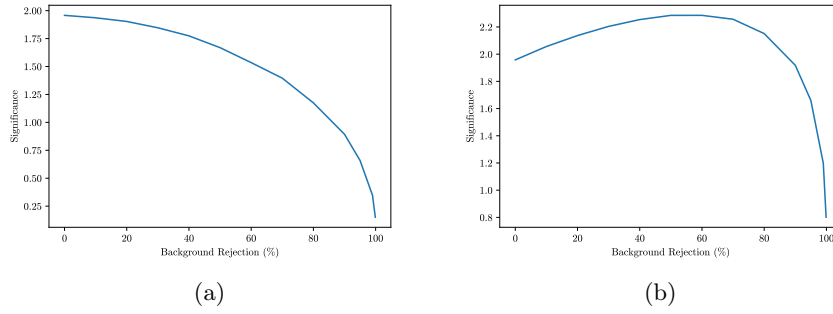


Figure D.2: DNN Response Distribution for the inclusive category for 300 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.1.3 500 GeV

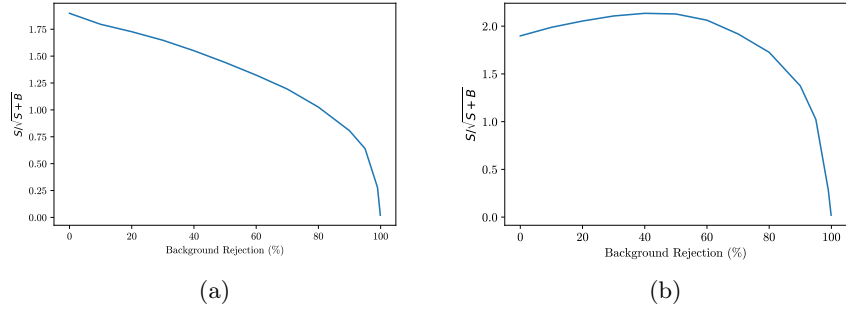


Figure D.3: DNN Response Distribution for the inclusive category for 500 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.2 $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$

D.2.1 150 GeV

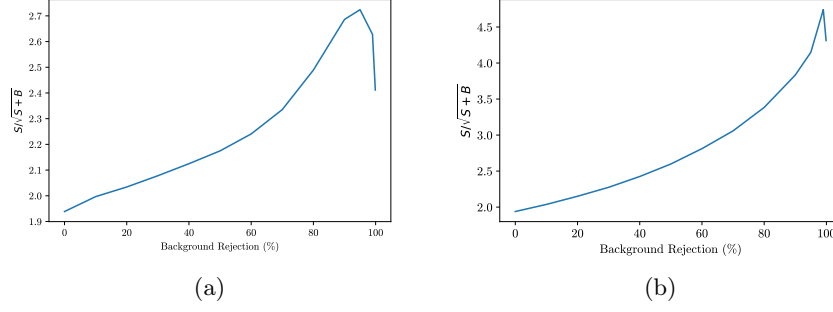


Figure D.4: DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.2.2 200 GeV

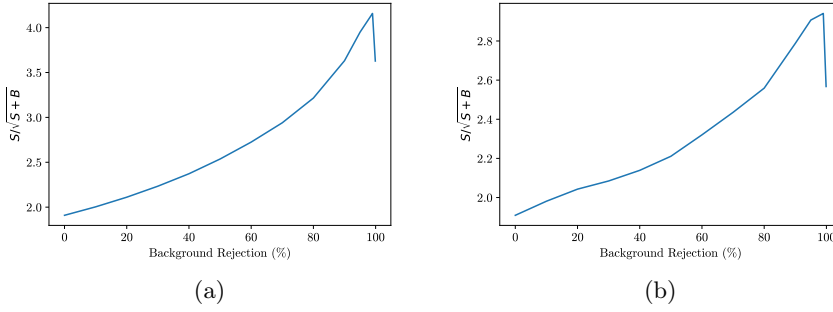


Figure D.5: DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Wh$ category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.3 $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$

D.3.1 150 GeV

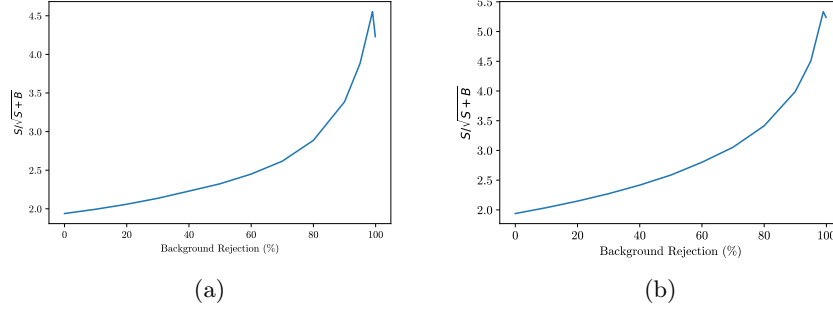


Figure D.6: DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.3.2 200 GeV

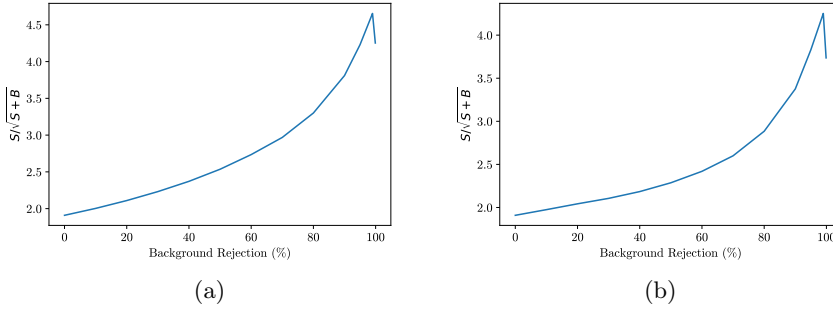


Figure D.7: DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.4 $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$

D.4.1 150 GeV

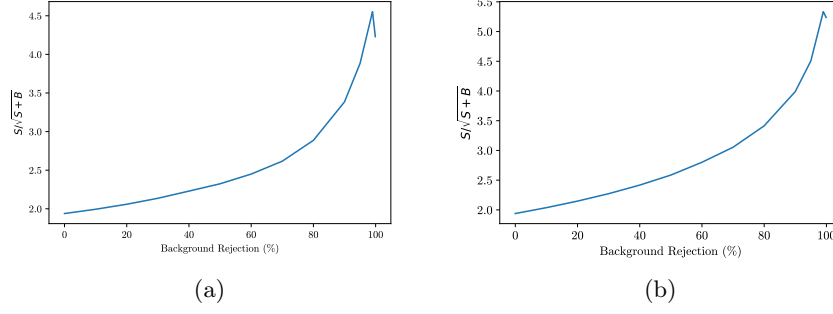


Figure D.8: DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.4.2 200 GeV

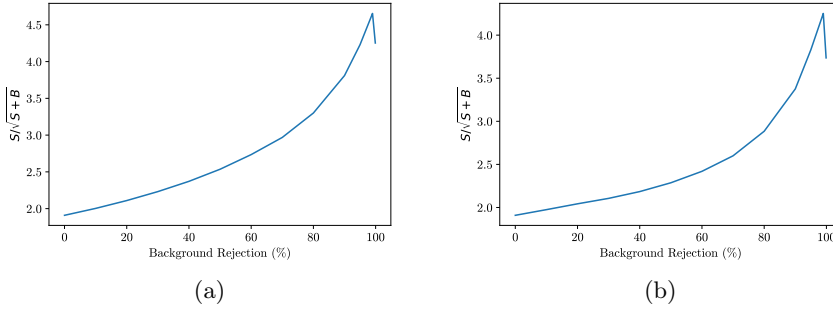


Figure D.9: DNN Response Distribution for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}, Zh$ category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.5 $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh

D.5.1 150 GeV

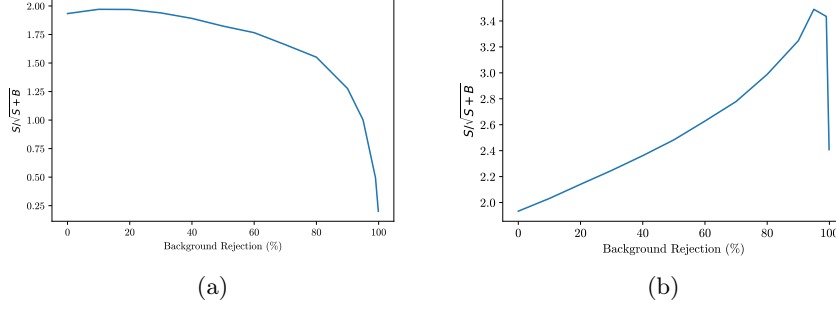


Figure D.10: DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.5.2 200 GeV

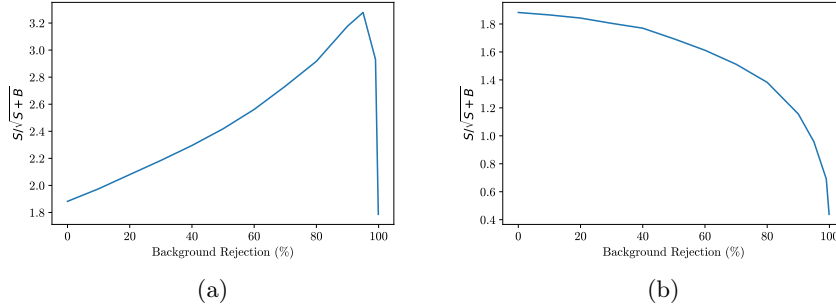


Figure D.11: DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Wh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.6 $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh

D.6.1 150 GeV

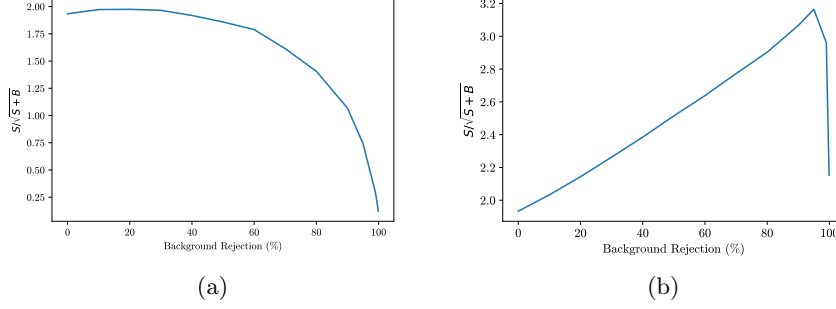


Figure D.12: DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.6.2 200 GeV

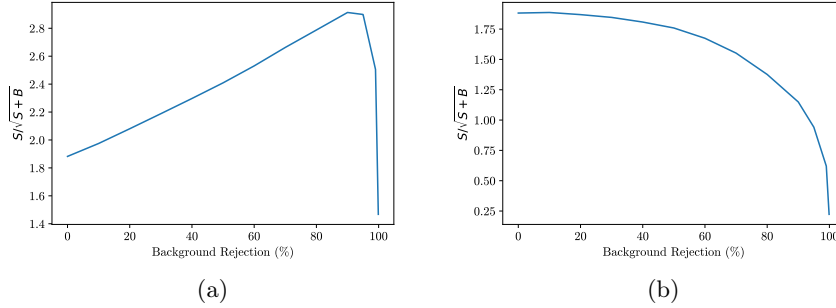


Figure D.13: DNN Response Distribution for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV, Zh category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.7 VBF

D.7.1 150 GeV

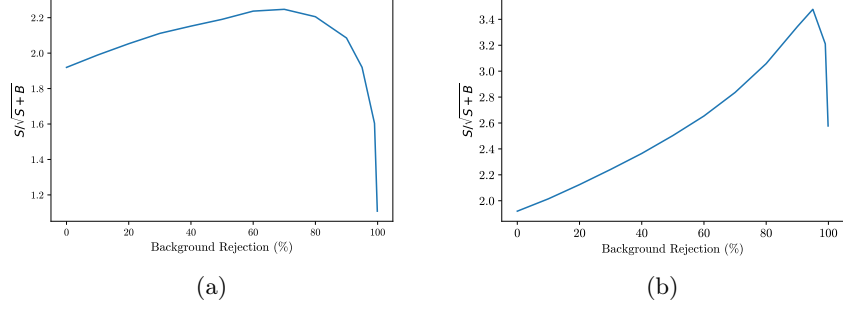


Figure D.14: DNN Response Distribution for the VBF category for 150 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.7.2 200 GeV

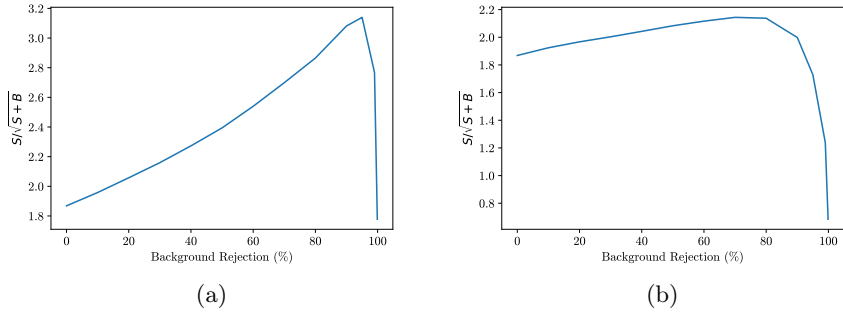


Figure D.15: DNN Response Distribution for the VBF category for 200 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.7.3 300 GeV

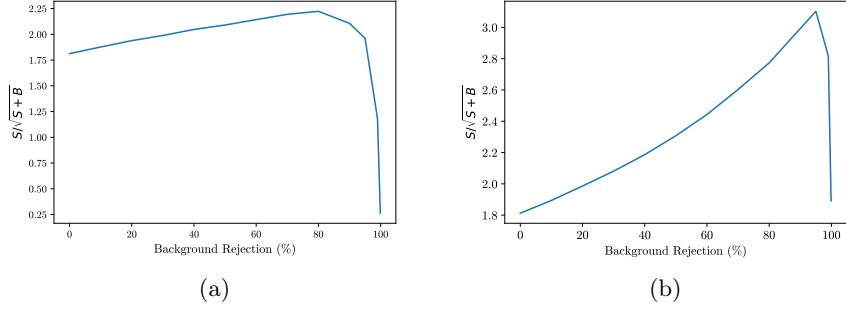


Figure D.16: Significance for the VBF category for 300 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

D.7.4 500 GeV

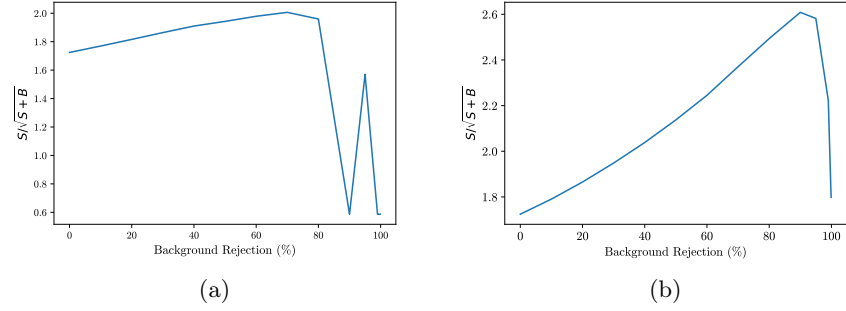


Figure D.17: DNN Response Distribution for the VBF category for 500 GeV mass point: (a) Weakly supervised learning and (b) Fully supervised learning.

Appendix E

ROC Curves

ROC Curve plots can be found in Appendix [E](#). The performance of the classifier is measured by the AUC, which is obtained by integrating these ROC curves. Figure [E.1](#) compares the performance of RF and DNN based on weakly supervised learning and fully supervised learning for the inclusive category for mass points 200, 300, and 500. The ROC curves for $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ and $N_{jets} \geq 2$; $60 < m_{jj} < 120 \text{ GeV}$, respectively, are shown in Figures [E.2](#) and [E.3](#). Figure [E.4](#) shows the VBF category.

E.1 Inclusive

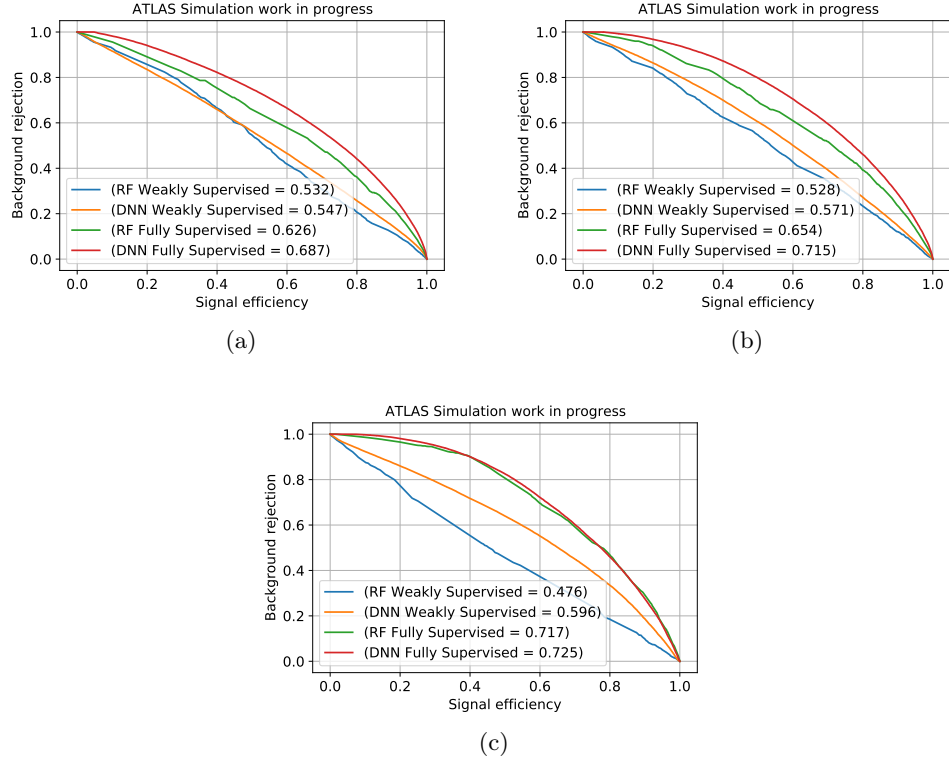


Figure E.1: ROC curves for the inclusive category for the following mass points: (a) 200 GeV, (b) 300 GeV and (c) 500 GeV learning.

E.2 $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$

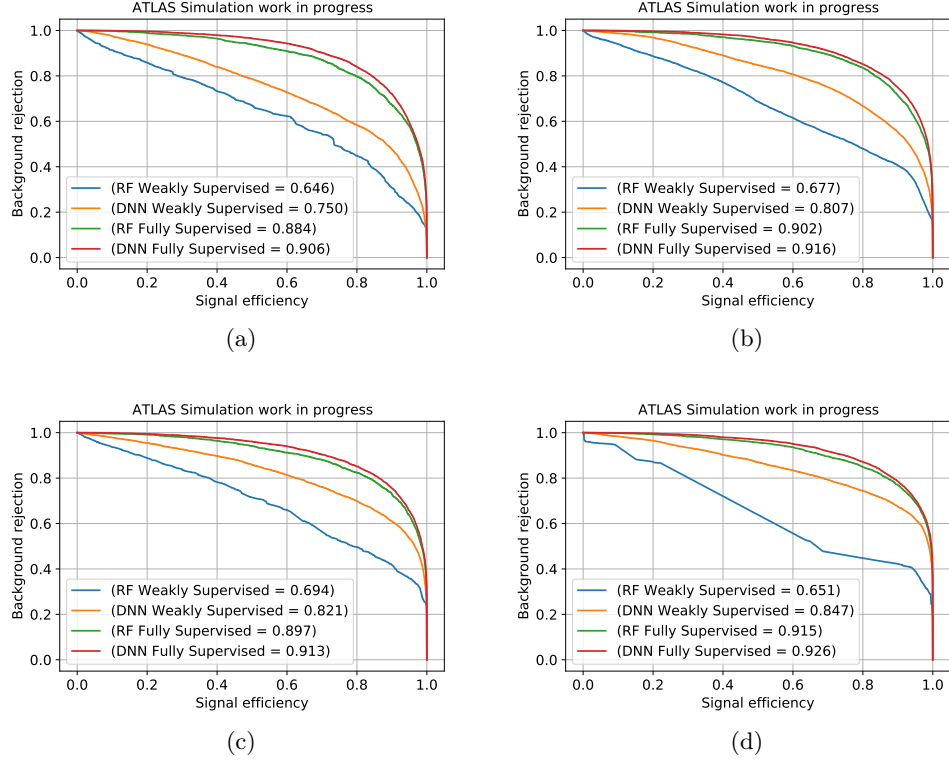


Figure E.2: ROC curves for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ category for the following mass points: (a) Wh 150 GeV, (b) Wh 200 GeV, (c) Zh 150 GeV and (d) Zh 200 GeV.

E.3 $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV category

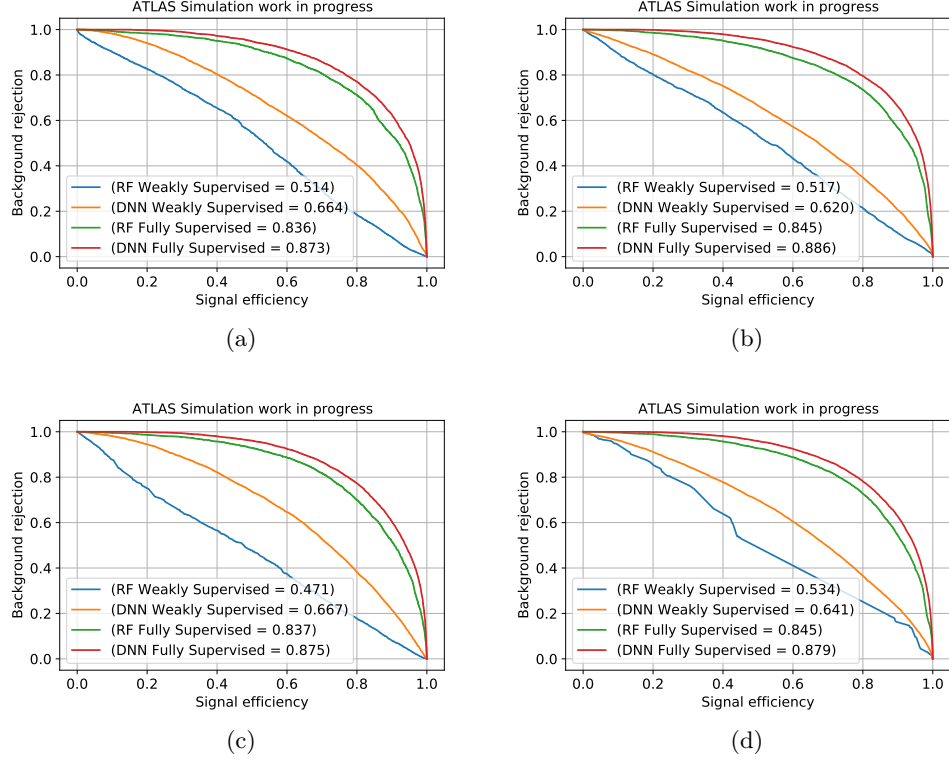


Figure E.3: ROC curves for the $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV category for the following mass points: (a) Wh 150 GeV, (b) Wh 200 GeV, (c) Zh 150 GeV and (d) Zh 200 GeV.

E.4 VBF Category

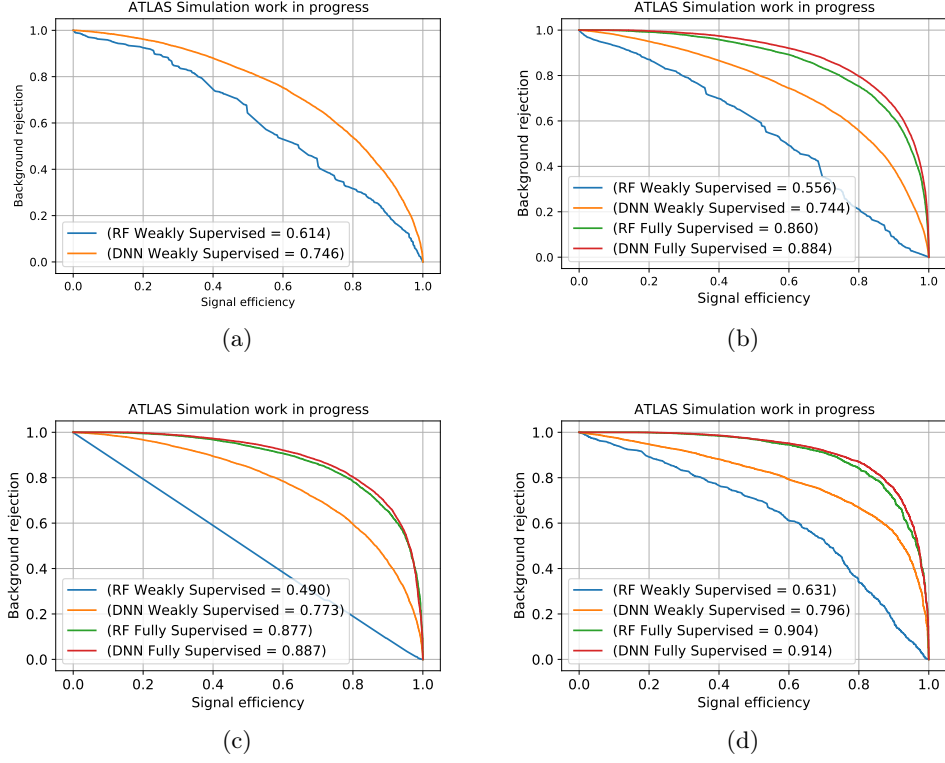


Figure E.4: ROC curves for the VBF category for the following mass points: (a) 150 GeV, (b) 200 GeV, (c) 300 GeV and (d) 500 GeV learning

Appendix F

Spurious Signal

The figures in this Appendix F show the study of the spurious signals for the other mass points for which fully simulated signal samples were available. The methodology used to collect the spurious signals is described in section 5.6. The fit of property of each tested function is presented in Table F.1, F.2 and F.3 for $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ $N_{jets} \geq 2, 75 \text{ GeV} \geq m_{jj} \geq 105 \text{ GeV}$ and $N_{jets} \geq 1$ categories, respectively. These Tables serve as a summary for Figures F.1 - F.2, F.3 - F.4 and F.5 - F.6, respectively.

F.1 $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$

Function	S/ Δ S in %	Spurious signal	N par	Chi2 Prob	Pass selection
Category-0					
FK0	72.5	43.7	2	28.7	PASS
FK1	178	148	3	24.2	PASS
Category-1					
FK0	62.2	35.3	2	24.4	PASS
FK1	163	122	3	19.1	PASS
Category-2					
FK0	63.2	-37.9	2	24.2	PASS
FK1	161	116	3	17.7	PASS
Category-3					
FK0	63.8	-31.4	2	23.5	PASS
FK1	153	92.6	3	18.7	PASS
Category-4					
FK0	-40.7	-25	2	43.3	PASS
FK1	108	72.2	3	36.7	PASS
Category-5					
FK0	46.4	-16.5	2	20.3	PASS
FK1	101	54.8	3	16.2	PASS
Category-6					
FK0	33.6	12.3	2	40.2	PASS
FK1	79	37.3	3	34.1	PASS

Table F.1: Fitted spurious signal yield at 150 GeV for VH Leptonic phase space, $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$.

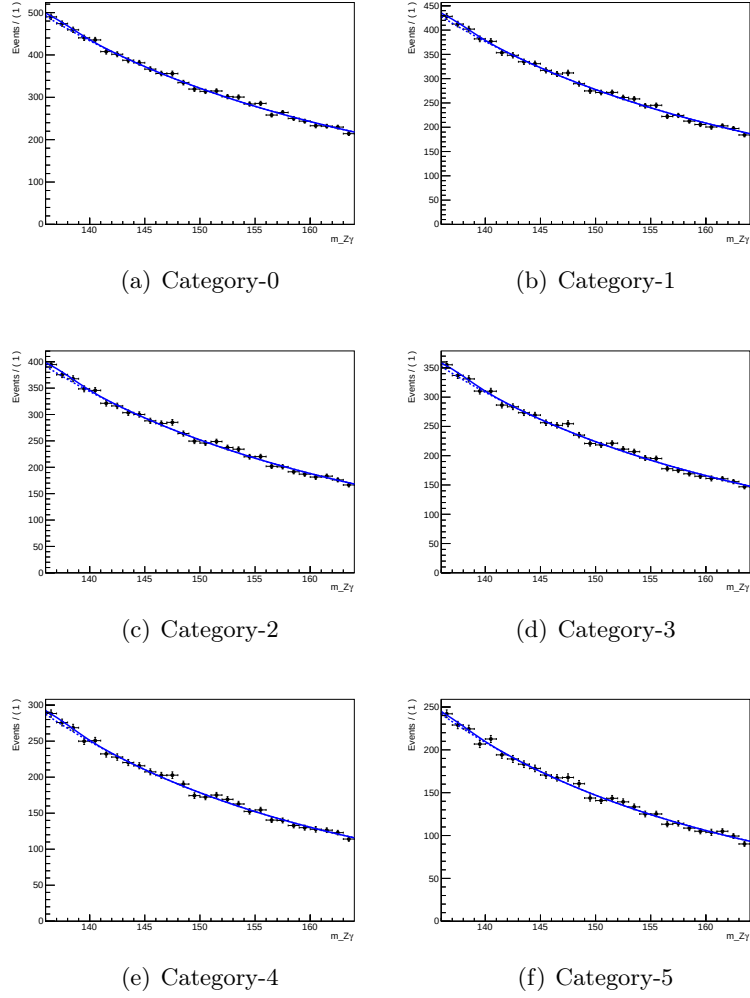


Figure F.1: Fitting on the background template for VH Leptonic 150 GeV, with S+B function.

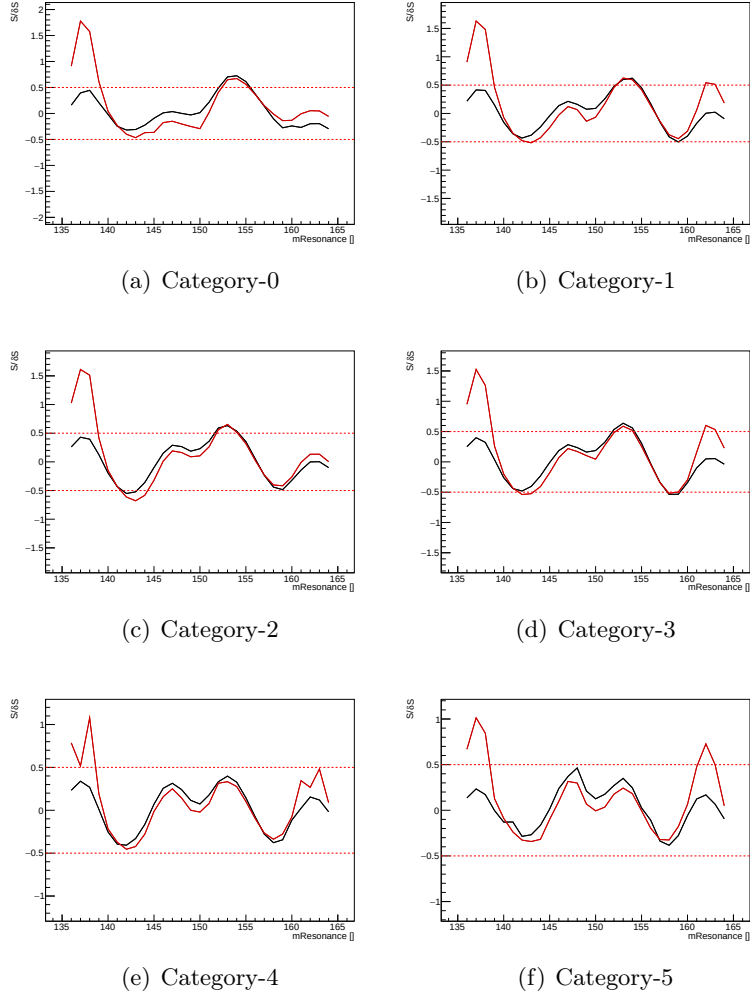


Figure F.2: The $N_{spurious}/\delta N_{spurious}$ as a function of mass for each functional form for VH Leptonic 150 GeV.

F.2 $N_{jets} \geq 2$ and $75 \leq m_{jj} \leq 105$ GeV

Function	S/ Δ S in %	Spurious signal	N par	Chi2 Prob	Pass selection
Category-0					
FK0	-32.6	-7.32	2	49.7	PASS
FK1	-52.8	16.7	3	47.4	PASS
Category-1					
FK0	-47.9	-7.56	2	64.2	PASS
FK1	-44.3	13.2	3	70.6	PASS
Category-2					
FK1	-39.6	8.84	3	83.6	PASS
FK0	-53.6	-10.4	2	69.2	PASS
Category-3					
FK0	-52.2	10.4	2	53.9	PASS
FK1	-59.7	-12	3	63.9	PASS
Category-4					
FK1	-26.1	-4.86	3	72.7	PASS
FK0	-68.9	-12.4	2	36.2	PASS
Category-5					
FK1	-31.8	5.85	3	85.2	PASS
FK0	-60.7	-9.94	2	57.9	PASS
Category-6					
FK0	54.5	9.42	2	42.1	PASS
FK1	66.9	11.9	3	45	PASS

Table F.2: Fitted spurious signal yield at 150 GeV for VH Hadronic phase space.

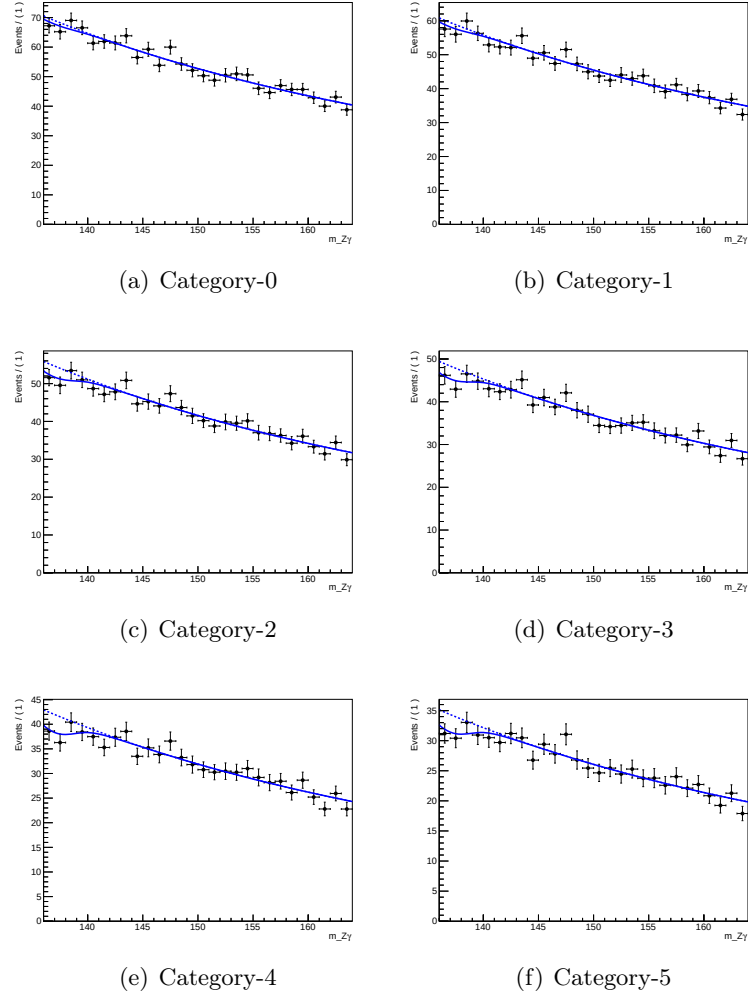


Figure F.3: Fitting on the background template for VH Hadronic 150 GeV, with S+B function.

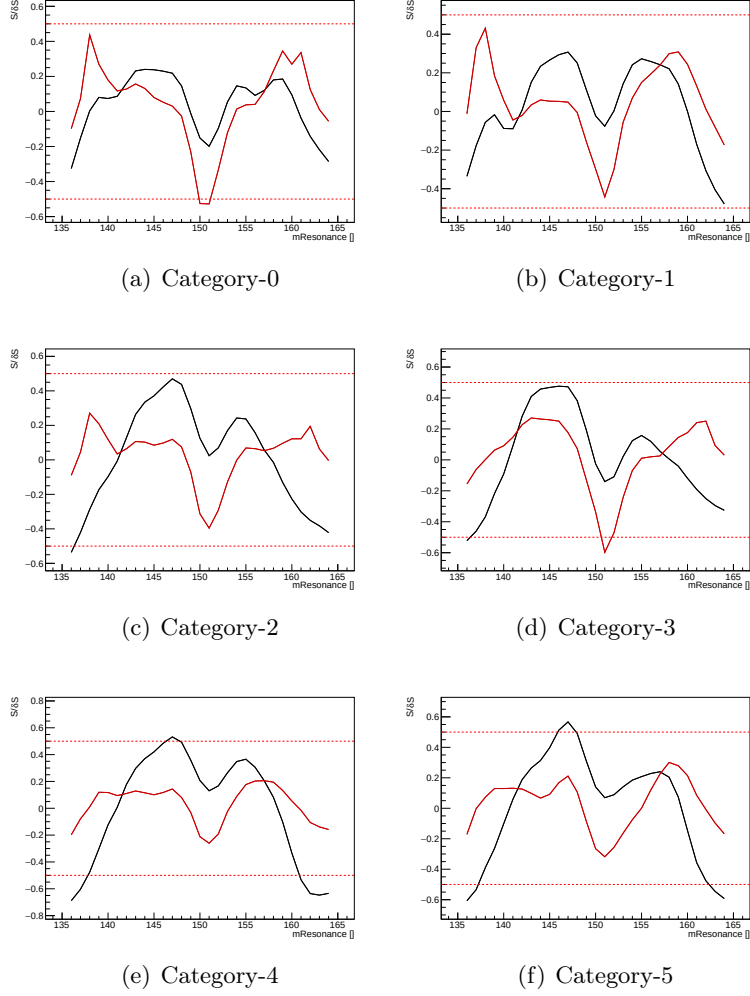


Figure F.4: The $N_{spurious}/\delta N_{spurious}$ as a function of mass for each functional form for VH Hadronic 150 GeV.

F.3 $N_{jets} \geq 1$

Function	S/ Δ S in %	Spurious signal	N par	Chi2 Prob	Pass selection
Category-0					
FK0	-32.6	-7.32	2	49.7	PASS
FK1	-52.8	16.7	3	47.4	PASS
Category-1					
FK0	-47.9	-7.56	2	64.2	PASS
FK1	-44.3	13.2	3	70.6	PASS
Category-2					
FK1	-39.6	8.84	3	83.6	PASS
FK0	-53.6	-10.4	2	69.2	PASS
Category-3					
FK0	-52.2	10.4	2	53.9	PASS
FK1	-59.7	-12	3	63.9	PASS
Category-4					
FK1	-26.1	-4.86	3	72.7	PASS
FK0	-68.9	-12.4	2	36.2	PASS
Category-5					
FK1	-31.8	5.85	3	85.2	PASS
FK0	-60.7	-9.94	2	57.9	PASS
Category-6					
FK0	54.5	9.42	2	42.1	PASS
FK1	66.9	11.9	3	45	PASS

Table F.3: Fitted spurious signal yield at 150 GeV for VH Leptonic phase space.

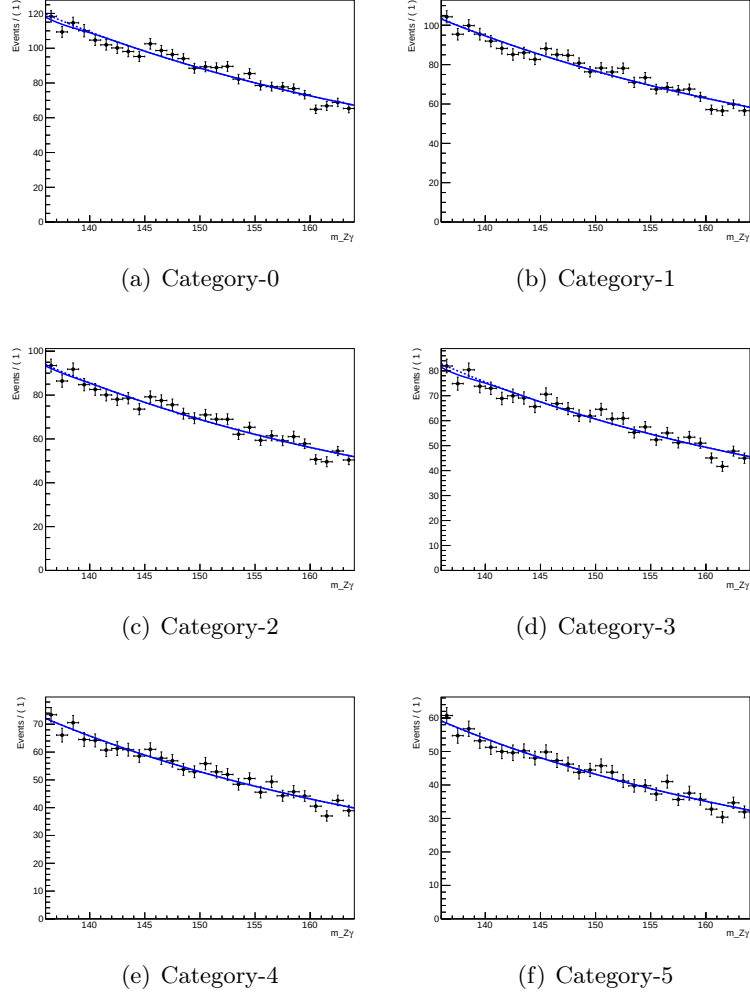


Figure F.5: Fitting on the background template for $t\bar{t}H$ 150 GeV, with S+B function.

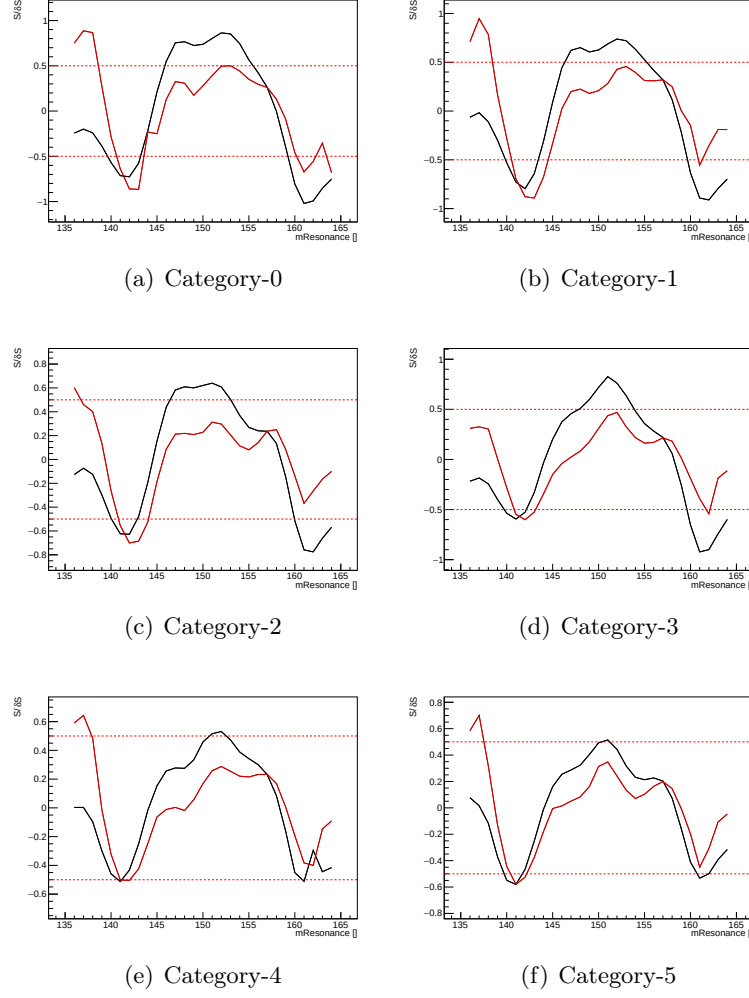


Figure F.6: The $N_{spurious}/\delta N_{spurious}$ as a function of mass for each functional form for ttH 150 GeV.

Appendix G

Signal Parameterization

The signal parameterization is performed on all the mass points for the categories considered for this study and the supplementary material is provided in Appendix G. The μ_{CB} and σ_{CB} for the $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ and $N_{jets} \geq 2$ and $75 \leq m_{jj} \leq 105 \text{ GeV}$ categories are summarized in Tables G.1 - G.4. These results are shown in Figure G.1 - G.4 as a function of the invariant mass.

Mass Point (GeV)	Background rejection (%)	μ_{CB} (GeV)	σ_{CB} (GeV)
150	10	149.74 ± 0.07	2.41 ± 0.08
	20	149.73 ± 0.07	2.42 ± 0.08
	30	149.73 ± 0.07	2.40 ± 0.08
	40	149.75 ± 0.07	2.40 ± 0.08
	50	149.76 ± 0.04	2.49 ± 0.05
200	10	199.73 ± 0.06	2.90 ± 0.07
	20	199.72 ± 0.06	2.93 ± 0.07
	30	199.72 ± 0.06	2.92 ± 0.07
	40	199.73 ± 0.06	2.92 ± 0.07
	50	199.77 ± 0.06	3.00 ± 0.06

Table G.1: Summary of μ_{CB} and σ_{CB} values for Wh 150 GeV signal sample at $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ after background rejection.

Mass Point (GeV)	Background rejection (%)	μ_{CB} (GeV)	σ_{CB} (GeV)
150	10	149.62 ± 0.07	2.34 ± 0.13
	20	149.59 ± 0.05	2.47 ± 0.04
	30	149.66 ± 0.05	2.56 ± 0.04
	40	149.67 ± 0.05	2.56 ± 0.04
	50	149.67 ± 0.06	2.55 ± 0.04
200	10	199.67 ± 0.88	2.92 ± 0.55
	20	199.69 ± 0.05	2.91 ± 0.06
	30	199.69 ± 0.04	2.91 ± 0.05
	40	199.59 ± 0.05	2.97 ± 0.04
	50	199.68 ± 0.05	2.90 ± 0.04

Table G.2: Summary of μ_{CB} and σ_{CB} values for Zh 150 GeV signal sample at $S_{ET}^{miss} > 2.5 \sqrt{\text{GeV}}$ after background rejection.

Background rejection (%)	μ_{CB} (GeV)	σ_{CB} (GeV)
150	10	149.62 ± 0.07
	20	149.63 ± 0.06
	30	149.62 ± 0.05
	40	149.69 ± 0.08
	50	$149.70 \pm \text{nan}$
200	10	199.64 ± 0.08
	20	199.56 ± 0.06
	30	199.51 ± 0.05
	40	199.61 ± 0.08
	50	$199.56 \pm \text{nan}$

Table G.3: Summary of μ_{CB} and σ_{CB} values for Wh 150 GeV signal sample at $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV after background rejection.

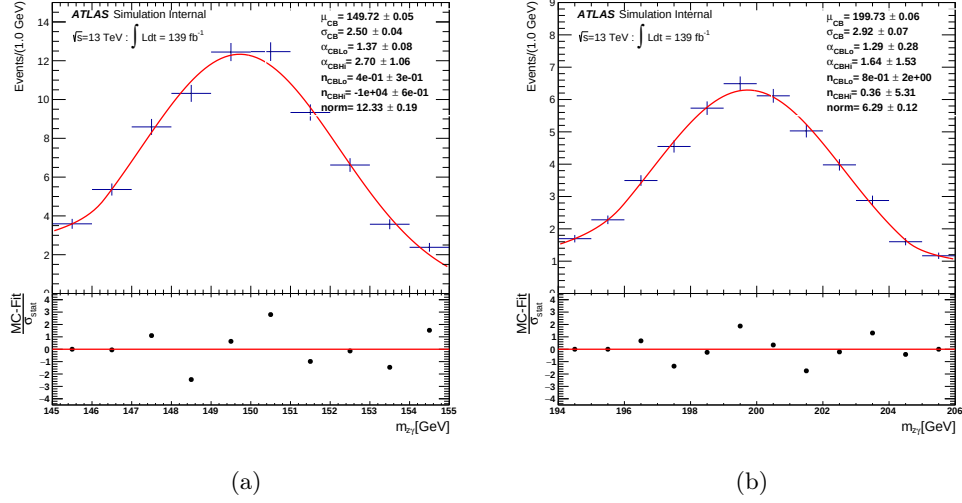


Figure G.1: Results of individual fit of DSCB parameters as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for Wh (metsig): (a) 150 GeV and (b) 200 GeV.

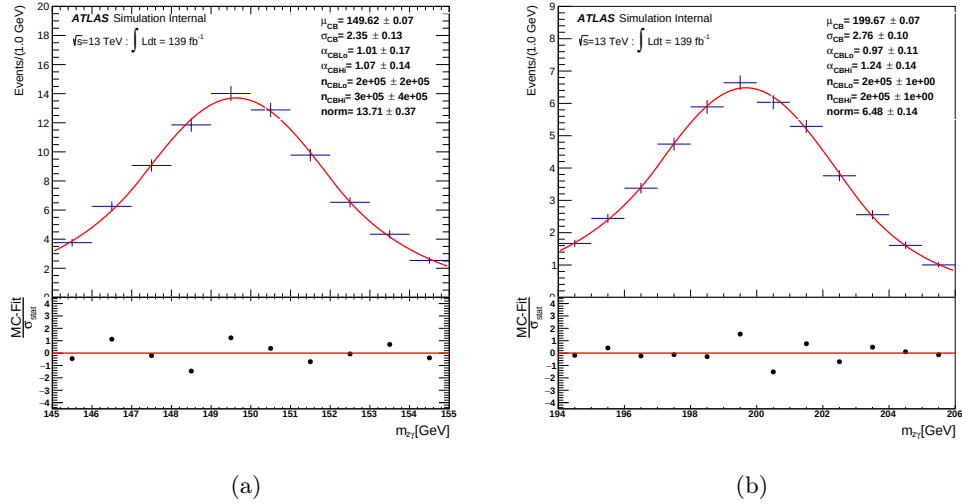


Figure G.2: Results of individual fit of DSCB parameters as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for Zh (metsig): (a) 150 GeV and (b) 200 GeV.

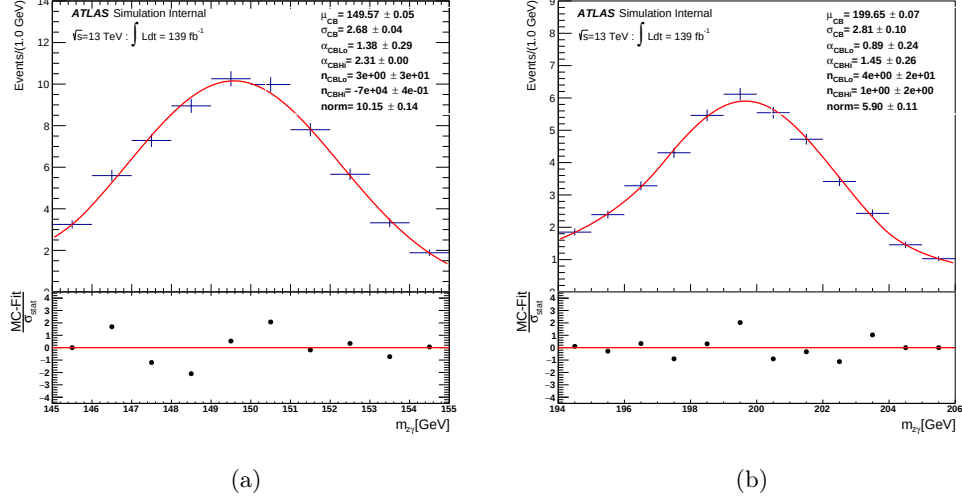
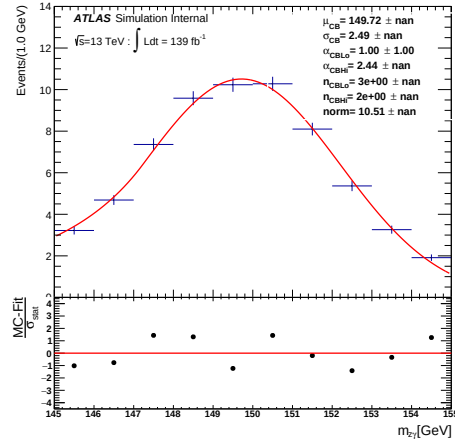


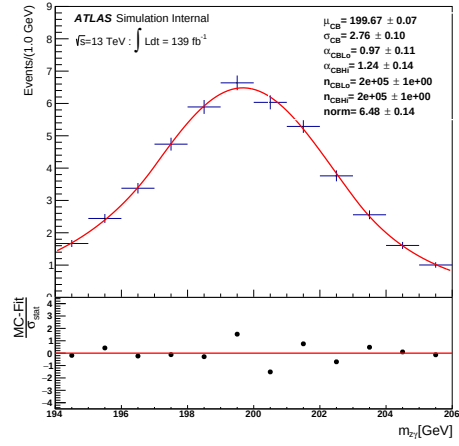
Figure G.3: Results of individual fit of DSCB parameters as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for $Wh(m_{jj})$: (a) 150 GeV and (b) 200 GeV.

Mass Point (GeV)	Background rejection (%)	μ_{CB} (GeV)	σ_{CB} (GeV)
150	10	149.66 ± 0.08	2.54 ± 0.07
	20	149.64 ± 0.05	2.45 ± 0.07
	30	149.61 ± 0.09	2.56 ± 0.08
	40	149.63 ± 0.05	2.48 ± 0.06
	50	149.63 ± 0.06	2.56 ± 0.05
200	10	199.63 ± 0.05	2.82 ± 0.09
	20	199.64 ± 0.05	2.82 ± 0.09
	30	199.67 ± 0.06	2.84 ± 0.09
	40	199.77 ± 1.00	3.01 ± 1.00
	50	199.73 ± 0.07	2.79 ± 0.13

Table G.4: Summary of μ_{CB} and σ_{CB} values for Zh 150 GeV signal sample at $N_{jets} \geq 2$ and $60 < m_{jj} < 120$ GeV after background rejection.



(a)



(b)

Figure G.4: Results of individual fit of DSCB parameters as a function of the invariant mass spectrum of $Z(\rightarrow \ell\ell)\gamma$ for $Zh(m_{jj})$: (a) 150 GeV and (b) 200 GeV.

Bibliography

- [1] S.-E. Dahbi, J. Choma, G. Mokgatitswane, X. Ruan, B. Lieberman, B. Mellado, and T. Celik, “Machine learning approach for the search of resonances with topological features at the Large Hadron Collider,” *International Journal of Modern Physics A*, vol. 37, no. 03, jan 2022. [Online]. Available: <https://doi.org/10.1142%2Fs0217751x21502419>
- [2] J. Choma, S.-E. Dahbi, B. Mellado, and X. Ruan, “The use of semi-supervision in the search for heavy resonances with the $Z\gamma$ final state at the LHC,” *The Proceedings of SAIP2021, the 65th Annual Conference of the South African Institute of Physics*, pp. 102 – 107, 2021.
- [3] S.-E. Dahbi, J. Choma, B. Mellado, and X. Ruan, “Machine learning approach for the search of resonances with topological features at the Large Hadron Collider,” *The Proceedings of SAIP2021, the 65th Annual Conference of the South African Institute of Physics*, pp. 163 – 168, 2021.
- [4] B. Lieberman, J. Choma, S.-E. Dahbi, B. Mellado, and X. Ruan, “An investigation of over-training within semi-supervised machine learning models in the search for heavy resonances at the LHC,” *arXiv preprint arXiv:2109.07287*, 2021.
- [5] G. Mokgatitswane, J. Choma, S.-E. Dahbi, X. Ruan, and B. Mellado, “Study of systematic uncertainties and spurious signals of resonant $H \rightarrow Z\gamma$ production at ATLAS experiment,” *The Proceedings of SAIP2022, the 66th Annual Conference of the South African Institute of Physics*, pp. 189 – 194, 2022.
- [6] S. von Buddenbrock, A. S. Cornell, Y. Fang, A. F. Mohammed, M. Kumar, B. Mellado, and K. G. Tomiwa, “The emergence of multi-lepton anomalies at the LHC and their compatibility with new physics at the EW scale,” *Journal of High Energy Physics*, vol. 2019, no. 10, pp. 1–41, 2019.
- [7] S. von Buddenbrock, N. Chakrabarty, A. S. Cornell, D. Kar, M. Kumar, T. Mandal, B. Mellado, B. Mukhopadhyaya, and R. G. Reed, “The compatibility of LHC Run 1 data with a heavy scalar of mass around 270 GeV,” 2015. [Online]. Available: <https://arxiv.org/abs/1506.00612>
- [8] S. von Buddenbrock, N. Chakrabarty, A. S. Cornell, D. Kar, M. Kumar, T. Mandal, B. Mellado, B. Mukhopadhyaya, R. G. Reed, and X. Ruan, “Phenomenological signatures of additional scalar bosons at the LHC,” *Eur. Phys. J. C*, vol. 76, no. 10, p. 580, 2016.

- [9] Y. Fang, M. Kumar, B. Mellado, Y. Zhang, and M. Zhu, “Impact of additional bosons on the exploration of the Higgs boson at the LHC,” *Int. J. Mod. Phys. A*, vol. 32, no. 34, p. 1746010, 2017.
- [10] S. von Buddenbrock, A. S. Cornell, A. Fadol, M. Kumar, B. Mellado, and X. Ruan, “Multi-lepton signatures of additional scalar bosons beyond the Standard Model at the LHC,” *J. Phys. G*, vol. 45, no. 11, p. 115003, 2018.
- [11] S. von Buddenbrock, A. S. Cornell, E. D. Iarilala, M. Kumar, B. Mellado, X. Ruan, and E. M. Shrif, “Constraints on a 2HDM with a singlet scalar and implications in the search for heavy bosons at the LHC,” *J. Phys. G*, vol. 46, no. 11, p. 115001, 2019.
- [12] D. Sabatta, A. S. Cornell, A. Goyal, M. Kumar, B. Mellado, and X. Ruan, “Connecting muon anomalous magnetic moment and multi-lepton anomalies at LHC,” *Chin. Phys. C*, vol. 44, no. 6, p. 063103, 2020.
- [13] Y. Hernandez, M. Kumar, A. S. Cornell, S.-E. Dahbi, Y. Fang, B. Lieberman, B. Mellado, K. Monnakgotla, X. Ruan, and S. Xin, “The anomalous production of multi-leptons and its impact on the measurement of Wh production at the LHC,” *The European Physical Journal C*, vol. 81, no. 4, p. 365, 2021.
- [14] S. von Buddenbrock, R. Ruiz, and B. Mellado, “Anatomy of inclusive $t\bar{t}w$ production at hadron colliders,” *Physics Letters B*, vol. 811, p. 135964, dec 2020. [Online]. Available: <https://doi.org/10.1016%2Fj.physletb.2020.135964>
- [15] O. Fischer, B. Mellado, S. Antusch, E. Bagnaschi, S. Banerjee, G. Beck, B. Belfatto, M. Bellis, Z. Berezhiani, M. Blanke, B. Capdevila, K. Cheung, A. Crivellin, N. Desai, B. Dev, R. Godbole, T. Han, P. Harris, M. Hoferichter, M. Kirk, S. Kulkarni, C. Lange, K. Lassila-Perini, Z. Liu, F. Mahmoudi, C. A. Manzari, D. Marzocca, B. Mukhopadhyaya, A. Pich, X. Ruan, L. Schnell, J. Thaler, and S. Westhoff, “Unveiling hidden physics at the LHC,” *The European Physical Journal C*, vol. 82, no. 8, aug 2022. [Online]. Available: <https://doi.org/10.1140%2Fepjc%2Fs10052-022-10541-4>
- [16] A. Crivellin, Y. Fang, O. Fischer, A. Kumar, M. Kumar, E. Malwa, B. Mellado, N. Rapheeha, X. Ruan, and Q. Sha, “Accumulating evidence for the associate production of a neutral scalar with mass around 151 GeV,” *arXiv preprint arXiv:2109.02650*, 2021.
- [17] P. B. Pal, *An introductory course of particle physics*. Taylor & Francis, 2014.
- [18] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012.
- [19] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Physics Letters B*, vol. 716, no. 1, pp. 30–61, 2012.

- [20] P. Langacker, *The Standard Model and beyond*. Taylor & Francis, 2017.
- [21] S. L. Glashow, “Partial-symmetries of weak interactions,” *Nuclear physics*, vol. 22, no. 4, pp. 579–588, 1961.
- [22] S. Weinberg, “A model of leptons,” *Physical review letters*, vol. 19, no. 21, p. 1264, 1967.
- [23] A. Salam, “Weak and electromagnetic interactions,” in *Selected Papers Of Abdus Salam: (With Commentary)*. World Scientific, 1994, pp. 244–254.
- [24] T. Shears, “The Standard Model,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 370, no. 1961, pp. 805–817, 2012.
- [25] D. Griffiths, *Introduction to elementary particles*. John Wiley & Sons, 2020.
- [26] M. Cush. Standard Model of elementary particles.svg - wikipedia, the free encyclopedia. [Online]. Available: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg
- [27] F. Halzen, A. Martin, and L. Quarks, “An introductory course in modern particle physics,” *John and Wiley*, 1984.
- [28] S. Riad, “Studies of effective theories beyond the Standard Model,” Ph.D. dissertation, KTH Royal Institute of Technology, 2014.
- [29] J. Goldstone, A. Salam, and S. Weinberg, “Broken symmetries,” *Physical Review*, vol. 127, no. 3, p. 965, 1962.
- [30] Alonso, Francisco and Anduaga, Xabier Sebastián and Dova, María Teresa and Monticelli, Fernando Gabriel and Tripiñana, Martín F, “A Particle Consistent with the Higgs Boson Observed with the ATLAS Detector at the Large Hadron Collider,” *Science*, vol. 338, p. 1576, 2012.
- [31] ATLAS Collaboration, “Measurement of the Higgs boson mass in the $H \rightarrow ZZ^* \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector,” *Physics Letters B*, vol. 784, pp. 345–366, 2018. [Online]. Available: <https://doi.org/10.1016%2Fj.physletb.2018.07.050>
- [32] J. Baglio and A. Djouadi, “Higgs production at the LHC,” *Journal of High Energy Physics*, vol. 2011, no. 3, pp. 1–53, 2011.
- [33] R. Tanaka. Higgs cross sections for HL-LHC and HE-LHC. [Online]. Available: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/HiggsEuropeanStrategy>
- [34] J. Brehmer, J. Jaeckel, and T. Plehn, “Polarized WW scattering on the Higgs pole,” *Physical Review D*, vol. 90, no. 5, p. 054023, 2014.
- [35] J. F. Gunion and H. E. Haber, “CP-conserving two-Higgs-doublet model: the approach to the decoupling limit,” *Physical Review D*, vol. 67, no. 7, p. 075019, 2003.

- [36] G. C. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, and J. P. Silva, “Theory and phenomenology of two-Higgs-doublet models,” *Physics reports*, vol. 516, no. 1-2, pp. 1–102, 2012.
- [37] Office of Science. DOE Explains the Standard Model of Particle Physics. [Online]. Available: <https://www.energy.gov/science/doe-explainsthe-standard-model-particle-physics>
- [38] S. von Buddenbrock, N. Chakrabarty, A. S. Cornell, D. Kar, M. Kumar, T. Mandal, B. Mellado, B. Mukhopadhyaya, R. G. Reed, and X. Ruan, “Phenomenological signatures of additional scalar bosons at the LHC,” *The European Physical Journal C*, vol. 76, no. 10, pp. 1–18, 2016.
- [39] S. von Buddenbrock, N. Chakrabartyb, A. S. Cornellc, D. Kara, M. Kumarc, T. Mandald, B. Melladoa, B. Mukhopadhyayab, R. G. Reeda, and X. Ruana, “Constraining hypothetical extensions to the Higgs sector at the LHC,” in *The Proceedings of SAIP2016, the 61st Annual Conference of the South African Institute of Physics*, S. Peterson and S. Yacoob, eds, pp. 158–163.
- [40] G. Aad, B. Abbott, J. Abdallah, S. A. Khalek, R. Aben, B. Abi, M. Abolins, O. AbouZeid, H. Abramowicz, H. Abreu *et al.*, “Fiducial and differential cross sections of Higgs boson production measured in the four-lepton decay channel in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector,” *Physics Letters B*, vol. 738, pp. 234–253, 2014.
- [41] CMS Collaboration, “Measurement of differential cross sections for Higgs boson production in the diphoton decay channel in pp collisions at $\sqrt{s} = 8$ TeV,” *The European Physical Journal C*, vol. 76, no. 1, pp. 1–31, 2016.
- [42] CMS Collaboration, “Measurement of differential and integrated fiducial cross sections for Higgs boson production in the four-lepton decay channel in pp collisions at $\sqrt{s}=7$ and 8 TeV,” *Journal of High Energy Physics*, vol. 2016, no. 4, pp. 1–46, 2016.
- [43] ATLAS Collaboration, “Search for the Standard Model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,” *Phys. Rev. D*, vol. 97, p. 072016, 2018.
- [44] CMS Collaboration, “Search for two Higgs bosons in final states containing two photons and two bottom quarks in proton-proton collisions at 8 TeV,” vol. 94, no. 5, p. 052012, 2016.
- [45] CMS Collaboration, “Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh , in the final states with $h \rightarrow \tau\tau$,” *Physics Letters B*, vol. 755, pp. 217–244, 2016.
- [46] CMS Collaboration, “Searches for heavy Higgs bosons in two-Higgs-doublet models and for $t \rightarrow c h$ decay using multilepton and diphoton final states in pp collisions at 8 TeV,” *Physical Review D*, vol. 90, no. 11, p. 112013, 2014.

- [47] ATLAS Collaboration, “Search for $H \rightarrow \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector,” *Physics Letters B*, vol. 740, pp. 222–242, 2015.
- [48] ATLAS Collaboration, “Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector,” *Physics Letters B*, vol. 749, pp. 519–541, oct 2015. [Online]. Available: <https://doi.org/10.1016%2Fj.physletb.2015.07.079>
- [49] ATLAS Collaboration, “Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s}=8$ TeV with the ATLAS detector,” *arXiv preprint arXiv:1503.05066*, 2015.
- [50] ATLAS Collaboration, “Search for the Standard Model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector,” *Physical Review D*, vol. 97, no. 7, p. 072016, 2018.
- [51] ATLAS Collaboration, “Search for a high-mass Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s}=8$ TeV with the ATLAS detector,” *Journal of High Energy Physics*, vol. 2016, no. 1, pp. 1–66, 2016.
- [52] ATLAS Collaboration, “Search for an additional, heavy Higgs boson in the $H \rightarrow ZZ$ decay channel at $\sqrt{s}=8$ TeV in pp collision data with the ATLAS detector,” *Eur. Phys. J. C*, vol. 76, p. 45, 2016.
- [53] CMS Collaboration, “Search for a Higgs boson in the mass range from 145 to 1000 GeV decaying to a pair of W or Z bosons,” *Journal of High Energy Physics*, vol. 2015, no. 10, oct 2015. [Online]. Available: <https://doi.org/10.1007%2Fjhep10%282015%29144>
- [54] S. von Buddenbrock, “Production of the Madala boson in association with top quarks,” *arXiv preprint arXiv:1901.08327*, 2019.
- [55] Stefan von Buddenbrock, “The construction of a scalar extension to the Standard Model and the search for a heavy scalar at $\sqrt{s}=13$ TeV with the ATLAS detector,” Ph.D. dissertation, The University of the Witwatersrand, 2020.
- [56] S. Von Buddenbrock, A. S. Cornell, A. F. Mohammed, M. Kumar, B. Mellado, and X. Ruan, “Multi-lepton signatures of additional scalar bosons beyond the Standard Model at the LHC,” *Journal of Physics G: Nuclear and Particle Physics*, vol. 45, no. 11, p. 115003, 2018.
- [57] L. R. Evans and P. Bryant, “LHC Machine,” *JINST*, vol. 3, p. S08001. 164 p, 2008, this report is an abridged version of the LHC Design Report (CERN-2004-003). [Online]. Available: <https://cds.cern.ch/record/1129806>
- [58] L. R. Evans, *The Large Hadron Collider: a marvel of technology*. Epfl Press, 2009.

- [59] D. Kar, *Experimental Particle Physics*. IOP Publishing Limited, 2019.
- [60] L. Evans and P. Bryant, “LHC machine, 2008.”
- [61] E. Mobs, “The CERN accelerator complex. Complexe des accélérateurs du CERN,” Jul 2016. [Online]. Available: <https://cds.cern.ch/record/2197559>
- [62] G. Aad, E. Abat, J. Abdallah, A. Abdelalim, A. Abdesselam, B. Abi, M. Abolins, H. Abramowicz, E. Acerbi, B. Acharya *et al.*, “The ATLAS experiment at the CERN Large Hadron Collider,” *Journal of instrumentation*, vol. 3, p. S08003, 2008.
- [63] CMS Collaboration, “The CMS experiment at the CERN LHC,” *JInst*, vol. 3, p. S08004, 2008.
- [64] LHCb Collaboration, “The LHCb detector at the LHC,” *Journal of instrumentation*, vol. 3, no. 08, pp. S08 005–S08 005, 2008.
- [65] K. Aamodt, A. A. Quintana, R. Achenbach, S. Acounis, D. Adamová, C. Adler, M. Aggarwal, F. Agnese, G. A. Rinella, Z. Ahammed *et al.*, “The ALICE experiment at the CERN LHC,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08002, 2008.
- [66] R. Hawkings. Luminosity public results Run2. [Online]. Available: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>
- [67] W. Herr and B. Muratori, “Concept of luminosity,” 2006.
- [68] K. Potamianos, “The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider,” *arXiv preprint arXiv:1608.07850*, 2016.
- [69] J. Pequeno. Computer generated image of the ATLAS Inner Detector. [Online]. Available: <https://cds.cern.ch/images/CERN-GE-0803014-01>
- [70] M. Capeans, G. Darbo, K. Einsweiler, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet, “ATLAS Insertable B-Layer Technical Design Report,” Tech. Rep., Sep 2010. [Online]. Available: <https://cds.cern.ch/record/1291633>
- [71] ATLAS Collaboration, “ATLAS Inner detector technical design report,” *Technical Design Report ATLAS. CERN, Geneva*, vol. 216, 1997.
- [72] A. Abdesselam, T. Akimoto, P. Allport, J. Alonso, B. Anderson, L. Andricek, F. Anghinolfi, R. Apsimon, G. Barbier, A. Barr *et al.*, “The barrel modules of the ATLAS semiconductor tracker,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 568, no. 2, pp. 642–671, 2006.
- [73] *ATLAS Inner Detector: Technical Design Report, 1*, ser. Technical design report. ATLAS. Geneva: CERN, 1997. [Online]. Available: <https://cds.cern.ch/record/331063>

- [74] S. Haywood, L. Rossi, R. Nickerson, and A. Romaniouk, *ATLAS Inner Detector: Technical Design Report, 2*, ser. Technical design report. ATLAS. Geneva: CERN, 1997. [Online]. Available: <https://cds.cern.ch/record/331064>
- [75] ATLAS Collaboration, “ATLAS Liquid-Argon calorimeter: Technical Design Report,” CERN-LHCC-96-41, Tech. Rep., 1996.
- [76] J. Abdallah, J. Proudfoot, S. Némécek, J. Pilcher, C. Vellidis, P. Lourtie, J. Li, A. Vartapetian, M. Simonyan, F. Marroquim *et al.*, “Design, construction and installation of the ATLAS hadronic barrel scintillator-tile calorimeter,” Tech. Rep., 2007.
- [77] A. Jelinskas, “Search for heavy neutral MSSM Higgs bosons in the di-tau final state with the ATLAS detector,” Ph.D. dissertation, University of Warwick, 2019.
- [78] S. Veneziano. ATLAS approved plots DAQ. (Date last accessed 02-January-2022). [Online]. Available: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ>
- [79] M. Aaboud, G. Aad, B. Abbott, J. Abdallah, B. Abeloos, R. Aben, O. AbouZeid, N. Abraham, H. Abramowicz, H. Abreu *et al.*, “Performance of the ATLAS trigger system in 2015,” *The European Physical Journal C*, vol. 77, no. 5, pp. 1–53, 2017.
- [80] J. Pequeno, “Event Cross Section in a computer generated image of the ATLAS detector,” Mar 2008. [Online]. Available: <https://cds.cern.ch/record/1096081>
- [81] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyses, and A. Salzburger, “Concepts, Design and Implementation of the ATLAS New Tracking (NEWT),” CERN, Geneva, Tech. Rep., Mar 2007, all figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2007-007>. [Online]. Available: <https://cds.cern.ch/record/1020106>
- [82] T. Cornelissen, M. Elsing, I. Gavrilenco, W. Liebig, E. Moyses, and A. Salzburger, “The new ATLAS track reconstruction (NEWT),” *Journal of Physics: Conference Series*, vol. 119, no. 3, p. 032014, jul 2008. [Online]. Available: <https://doi.org/10.1088/1742-6596/119/3/032014>
- [83] ATLAS Collaboration, “Performance of the ATLAS Inner Detector track and vertex reconstruction in the high pile-up LHC environment.” ATLAS-CONF-2012-042, 2012.
- [84] ATLAS Collaboration, “A neural network clustering algorithm for the ATLAS silicon pixel detector,” 2014.
- [85] ATLAS Collaboration, “Electron efficiency measurements with the ATLAS detector using 2012 LHC proton-proton collision data,” *arXiv preprint arXiv:1612.01456*, 2016.

- [86] W. Lampl, P. Loch, S. Menke, S. Rajagopalan, S. Laplace, G. Unal, H. Ma, S. Snyder, D. Lelas, and D. Rousseau, “Calorimeter clustering algorithms: description and performance,” Tech. Rep., 2008.
- [87] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-kt jet clustering algorithm,” *Journal of High Energy Physics*, vol. 2008, no. 04, p. 063, 2008.
- [88] ATLAS Collaboration, “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}=13$ TeV,” *The European Physical Journal C*, vol. 76, no. 5, pp. 1–30, 2016.
- [89] ATLAS Collaboration, “Measurement of Higgs boson production in the diphoton decay channel in p p collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector,” *Physical Review D*, vol. 90, no. 11, p. 112015, 2014.
- [90] F. A. Thiele, *An ATLAS Search for Sterile Neutrinos*. Niels Bohr Institute, Copenhagen University, 2019.
- [91] ATLAS Collaboration, “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s}=13$ TeV,” *arXiv preprint arXiv:1802.08168*, 2018.
- [92] T. Cohen, M. Freytsis, and B. Ostdiek, “(machine) learning to do more with less,” *Journal of High Energy Physics*, vol. 2018, no. 2, p. 34, 2018.
- [93] K. Albertsson, P. Altoe, D. Anderson, M. Andrews, J. P. A. Espinosa, A. Aurisano, L. Basara, A. Bevan, W. Bhimji, D. Bonacorsi *et al.*, “Machine learning in high energy physics community white paper,” in *Journal of Physics: Conference Series*, vol. 1085, no. 2. IOP Publishing, 2018, p. 022008.
- [94] P. C. Bhat, “Multivariate analysis methods in particle physics,” *Annual Review of Nuclear and Particle Science*, vol. 61, pp. 281–309, 2011.
- [95] F. Carminati, G. Khattak, M. Pierini, S. Vallecorsa, and A. Farbin, “Calorimetry with deep learning: particle classification, energy regression, and simulation for high-energy physics,” in *NIPS*, 2017.
- [96] CMS Collaboration, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s}=8$ TeV,” *arXiv preprint arXiv:1502.02701*, 2015.
- [97] CMS Collaboration, “Reconstruction and identification of τ lepton decays to hadrons and $\nu\tau$ at CMS,” 2016.
- [98] I. Narsky and F. C. Porter, *Statistical analysis techniques in particle physics: Fits, density estimation and supervised learning*. John Wiley & Sons, 2013.
- [99] R. M. Leone, “Machine learning multi-stage classification and regression in the search for vector-like quarks and the neyman construction in signal searches,” Ph.D. dissertation, The University of Arizona, 2016.

- [100] ATLAS Collaboration, “Search for vectorlike B quarks in events with one isolated lepton, missing transverse momentum, and jets at $\sqrt{s} = 8$ TeV with the ATLAS detector,” *Phys. Rev. D*, vol. 91, p. 112011, 2015.
- [101] P. J. Clark, V. J. Martin, A. Collaboration *et al.*, “Search for a multi-higgs-boson cascade in $W^+W^-b\bar{b}$ events with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV,” *Physical Review*, vol. 89, no. 3, p. 032002, 2014.
- [102] C. P. Lampen, *Search for a three lepton and missing transverse energy signature of supersymmetry using a multivariate analysis*. The University of Arizona, 2013.
- [103] S. Vazquez *et al.*, “Measurement of the associated production of a vector boson (W, Z) and top quark pair in the opposite sign dilepton channel with pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector,” 2014.
- [104] A. Zucchetta, “Searches for signatures of an extended Higgs sector in final states with leptons and Higgs to $b\bar{b}$ decays at CMS,” Ph.D. dissertation, Padua U., 2015.
- [105] ATLAS Collaboration, “Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated (W/Z) H production with the ATLAS detector,” *arXiv preprint arXiv:1409.6212*, 2014.
- [106] ATLAS Collaboration, “Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector,” *Journal of high energy physics*, vol. 2015, no. 4, pp. 1–74, 2015.
- [107] J. Stevens and M. Williams, “uboot: A boosting method for producing uniform selection efficiencies from multivariate classifiers,” *Journal of Instrumentation*, vol. 8, no. 12, p. P12013, 2013.
- [108] H. Esch, “Measurement of the top quark mass in topologies enhanced with single top-quarks produced in the t -channel using flavour-tagging and a neural network with ATLAS data at $\sqrt{s} = 8$ TeV,” Ph.D. dissertation, Dortmund U., 2015.
- [109] Y. Abulaiti, H. Åkerstedt, K. Bendtz, G. Bertoli, O. Bessidskaia, C. Bohm, C. Clément, W. A. Cribbs, D. Eriksson, K. Gellerstedt *et al.*, “Comprehensive measurements of t -channel single top-quark production cross sections at $\sqrt{s} = 7$ TeV with the ATLAS detector,” *Physical Review D*, vol. 90, no. 11, p. 112006, 2014.
- [110] A. Kalogeropoulos, “Search for direct stop pair production at the LHC with the CMS experiment,” Ph.D. dissertation, DESY, 2013.
- [111] J. Zhong, R.-S. Huang, and S.-C. Lee, “A program for the bayesian neural network in the root framework,” *Computer Physics Communications*, vol. 182, no. 12, pp. 2655–2660, 2011.
- [112] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, “Boosted decision trees as an alternative to artificial neural networks for particle identification,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 543, no. 2-3, pp. 577–584, 2005.

- [113] R. Aaij, J. Albrecht, F. Alessio, S. Amato, E. Aslanides, I. Belyaev, M. Van Beuzekom, E. Bonaccorsi, R. Bonnefoy, L. Brarda *et al.*, “The LHCb trigger and its performance in 2011,” *Journal of Instrumentation*, vol. 8, no. 04, p. P04022, 2013.
- [114] V. V. Gligorov and M. Williams, “Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree,” *Journal of Instrumentation*, vol. 8, no. 02, p. P02013, 2013.
- [115] M. Tosi, “Performance of tracking, b-tagging and jet/met reconstruction at the CMS high level trigger,” in *Journal of Physics: Conference Series*, vol. 664, no. 8. IOP Publishing, 2015, p. 082055.
- [116] CMS Collaboration, “Performance of tau-lepton reconstruction and identification in CMS,” *Journal of Instrumentation*, vol. 7, 2012.
- [117] J. Franc, P. Bouř, M. Štěpánek, and V. Kůs, “New statistical techniques in the measurement of the inclusive top pair production cross section,” 2014. [Online]. Available: <https://arxiv.org/abs/1412.1076>
- [118] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [119] W.-Z. Lin, J.-A. Fang, X. Xiao, and K.-C. Chou, “idna-prot: identification of dna binding proteins using random forest with grey model,” *PloS one*, vol. 6, no. 9, p. e24756, 2011.
- [120] R. Casanova, S. Saldana, E. Y. Chew, R. P. Danis, C. M. Greven, and W. T. Ambrosius, “Application of random forests methods to diabetic retinopathy classification analyses,” *PLOS one*, vol. 9, no. 6, p. e98587, 2014.
- [121] S. Hartini, Z. Rustam, G. S. Saragih, and M. J. S. Vargas, “Estimating probability of banking crises using random forest,” *IAES International Journal of Artificial Intelligence*, vol. 10, no. 2, p. 407, 2021.
- [122] L. Hu, J. Chen, J. Vaughan, S. Aramideh, H. Yang, K. Wang, A. Sudjianto, and V. N. Nair, “Supervised machine learning techniques: An overview with applications to banking,” *International Statistical Review*, vol. 89, no. 3, pp. 573–604, 2021.
- [123] X. Niu, C. Li, and X. Yu, “Predictive analytics of e-commerce search behavior for conversion,” 2017.
- [124] R. Joshi, R. Gupte, P. Saravanan *et al.*, “A random forest approach for predicting online buying behavior of indian customers,” *Theoretical Economics Letters*, vol. 8, no. 03, p. 448, 2018.
- [125] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.

- [126] M. Ghosh and P. Prabu, “Empirical analysis of ensemble methods for the classification of robocalls in telecommunications,” *International Journal of Electrical and Computer Engineering*, vol. 9, no. 4, pp. 3108–3114, 2019.
- [127] T. Trzeciński, Ł. Graczykowski, M. Glinka, A. Collaboration *et al.*, “Using random forest classifier for particle identification in the ALICE Experiment,” in *Conference on Information Technology, Systems Research and Computational Physics*. Springer, 2018, pp. 3–17.
- [128] T. Trzcinski and K. Deja, “Assigning quality labels in the high-energy physics experiment ALICE using machine learning algorithms,” *Proceedings of NICA days*, vol. 647, p. 655, 2017.
- [129] V. M. Abazov, B. Abbott, B. Acharya, M. Adams, T. Adams, G. Alexeev, G. Alkhazov, A. Alton, G. Alverson, G. Alves *et al.*, “Measurement of the top quark pair production cross section in the lepton+ jets channel in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV,” *Physical Review D*, vol. 84, no. 1, p. 012008, 2011.
- [130] Abazov, Victor Mukhamedovich and Abbott, B and Acharya, BS and Adams, M and Adams, T and Alexeev, GD and Alkhazov, G and Alton, A and Alverson, G and Alves, GA and others, “Measurements of WW and WZ production in W + jets final states in $p p^-$ collisions,” *Physical review letters*, vol. 108, no. 18, p. 181803, 2012.
- [131] V. M. Abazov, B. Abbott, B. Acharya, M. Adams, T. Adams, G. Alexeev, G. Alkhazov, A. Alton, A. Askew, S. Atkins *et al.*, “Combined search for the Higgs boson with the d0 experiment,” *Physical Review D*, vol. 88, no. 5, p. 052011, 2013.
- [132] V. M. Abazov, B. Abbott, B. Acharya, M. Adams, T. Adams, J. Agnew, G. Alexeev, G. Alkhazov, A. Alton, A. Askew *et al.*, “Improved b quark jet identification at the d0 experiment,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 763, pp. 290–303, 2014.
- [133] J. Lees, V. Poireau, V. Tisserand, J. G. Tico, E. Grauges, A. Palano, G. Eigen, B. Stugu, D. N. Brown, L. Kerth *et al.*, “Exclusive measurements of $b \rightarrow s \gamma$ transition rate and photon energy spectrum,” *Physical Review D*, vol. 86, no. 5, p. 052012, 2012.
- [134] J. Lees, V. Poireau, V. Tisserand, J. G. Tico, E. Grauges, A. Palano, G. Eigen, B. Stugu, D. Brown, L. Kerth *et al.*, “Search for di-muon decays of a low-mass Higgs boson in radiative decays of the γ (1 s),” *Physical Review D*, vol. 87, no. 3, p. 031102, 2013.
- [135] K. Lasocha, E. Richter-Was, D. Tracz, Z. Was, and P. Winkowska, “Machine learning classification: Case of Higgs boson C P state in $H \rightarrow \tau \tau$ decay at the LHC,” *Physical Review D*, vol. 100, no. 11, p. 113001, 2019.
- [136] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

- [137] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [138] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [139] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.
- [140] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.
- [141] A. Graves and N. Jaitly, “Towards end-to-end speech recognition with recurrent neural networks,” in *International conference on machine learning*. PMLR, 2014, pp. 1764–1772.
- [142] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in *Advances in neural information processing systems*, 2014, pp. 3104–3112.
- [143] Y. Bengio, “Learning deep architectures for ai,” *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- [144] D. Guest, K. Cranmer, and D. Whiteson, “Deep learning and its application to lhc physics,” *Annual Review of Nuclear and Particle Science*, vol. 68, pp. 161–181, 2018.
- [145] P. Baldi, P. Sadowski, and D. Whiteson, “Searching for exotic particles in high-energy physics with deep learning,” *Nature communications*, vol. 5, no. 1, pp. 1–9, 2014.
- [146] M. Bianchini and F. Scarselli, “On the complexity of shallow and deep neural network classifiers.”
- [147] C. Szegedy, A. Toshev, and D. Erhan, “Deep neural networks for object detection,” in *Advances in Neural Information Processing Systems*, 2013, pp. 2553–2561.
- [148] M. Abdughani, J. Ren, L. Wu, J.-M. Yang, and J. Zhao, “Supervised deep learning in high energy phenomenology: a mini review,” *Communications in Theoretical Physics*, vol. 71, no. 8, p. 955, 2019.
- [149] P. Baldi, P. Sadowski, and D. Whiteson, “Enhanced higgs to $\tau\tau$ -search with deep learning.”
- [150] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, “Deep learning in color: towards automated quark/gluon jet discrimination,” *Journal of High Energy Physics*, vol. 2017, no. 1, pp. 1–23, 2017.

- [151] T. Cheng, “Recursive neural networks in quark/gluon tagging,” *Computing and Software for Big Science*, vol. 2, no. 1, pp. 1–13, 2018.
- [152] V. Abazov, B. Abbott, A. Abdesselam, M. Abolins, V. Abramov, B. Acharya, D. Adams, M. Adams, S. Ahmed, G. Alexeev *et al.*, “Search for single top quark production at $d\bar{o}$ using neural networks,” *Physics Letters B*, vol. 517, no. 3-4, pp. 282–294, 2001.
- [153] D. Acosta, J. Adelman, T. Affolder, T. Akimoto, M. Albrow, D. Ambrose, S. Amerio, D. Amidei, A. Anastassov, K. Anikeev *et al.*, “Measurement of the cross section for $t\bar{t}$ production in $p\bar{p}$ collisions using the kinematics of lepton+ jets events,” *Physical Review D*, vol. 72, no. 5, p. 052003, 2005.
- [154] D. Whiteson and N. Naumann, “Support vector regression as a signal discriminator in high energy physics,” *Neurocomputing*, vol. 55, no. 1-2, pp. 251–264, 2003.
- [155] S. Whiteson and D. Whiteson, “Machine learning for event selection in high energy physics,” *Engineering Applications of Artificial Intelligence*, vol. 22, no. 8, pp. 1203–1217, 2009.
- [156] L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman, “Weakly supervised classification in high energy physics,” *Journal of High Energy Physics*, vol. 2017, no. 5, pp. 1–11, 2017.
- [157] S. Cheong, “Cs 229 project report: Weakly supervised classifiers with adversarial training in high-energy physics.”
- [158] C. Fyfe, “Artificial neural networks and information theory,” *Department of Computing and Information System, The university of Paisley*, 2000.
- [159] J. Choma, “Ensembles of neural networks for time series with application to climate change prediction,” Master’s thesis, 2018.
- [160] O. Chapelle, M. Chi, and A. Zien, “A continuation method for semi-supervised svms,” in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 185–192.
- [161] Y. Luo, J. Zhu, M. Li, Y. Ren, and B. Zhang, “Smooth neighbors on teacher graphs for semi-supervised learning,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8896–8905.
- [162] N. Al-Azzam and I. Shatnawi, “Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer,” *Annals of Medicine and Surgery*, vol. 62, pp. 53–64, 2021.
- [163] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, “Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3325–3337, 2014.

- [164] T. Ishida, G. Niu, and M. Sugiyama, “Binary classification from positive-confidence data,” *arXiv preprint arXiv:1710.07138*, 2017.
- [165] G. Hartmann, M. Grundmann, J. Hoffman, D. Tsai, V. Kwatra, O. Madani, S. Vijayanarasimhan, I. Essa, J. Rehg, and R. Sukthankar, “Weakly supervised learning of object segmentations from web-scale video,” in *European Conference on Computer Vision*. Springer, 2012, pp. 198–208.
- [166] P. Siva and T. Xiang, “Weakly supervised object detector learning with model drift detection,” in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 343–350.
- [167] F. Zhang, B. Du, L. Zhang, and M. Xu, “Weakly supervised learning based on coupled convolutional neural networks for aircraft detection,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5553–5563, 2016.
- [168] M. Dehghani, H. Zamani, A. Severyn, J. Kamps, and W. B. Croft, “Neural ranking models with weak supervision,” in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017, pp. 65–74.
- [169] D. J. Crandall and D. P. Huttenlocher, “Weakly supervised learning of part-based spatial models for visual object recognition,” in *European conference on computer vision*. Springer, 2006, pp. 16–29.
- [170] Y. Wang, S. Sohn, S. Liu, F. Shen, L. Wang, E. J. Atkinson, S. Amin, and H. Liu, “A clinical text classification paradigm using weak supervision and deep representation,” *BMC medical informatics and decision making*, vol. 19, no. 1, p. 1, 2019.
- [171] G. Lee, J. Jeong, S. Seo, C. Kim, and P. Kang, “Sentiment classification with word localization based on weakly supervised learning with a convolutional neural network,” *Knowledge-Based Systems*, vol. 152, pp. 70–82, 2018.
- [172] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, “Is object localization for free?-weakly-supervised learning with convolutional neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 685–694.
- [173] Z.-H. Zhou, “A brief introduction to weakly supervised learning,” *National Science Review*, vol. 5, no. 1, pp. 44–53, 2018.
- [174] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, “Semantic annotation of high-resolution satellite images via weakly supervised learning,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3660–3671, 2016.
- [175] H. Kuehne, A. Richard, and J. Gall, “Weakly supervised learning of actions from transcripts,” *Computer Vision and Image Understanding*, vol. 163, pp. 78–89, 2017.

- [176] D.-A. Huang, L. Fei-Fei, and J. C. Niebles, “Connectionist temporal modeling for weakly supervised action labeling,” in *European Conference on Computer Vision*. Springer, 2016, pp. 137–153.
- [177] J. Hernández-González, I. Inza, and J. A. Lozano, “Weak supervision and other non-standard classification problems: a taxonomy,” *Pattern Recognition Letters*, vol. 69, pp. 49–55, 2016.
- [178] M. Dehghani, A. Severyn, S. Rothe, and J. Kamps, “Learning to learn from weak supervision by full supervision,” *arXiv preprint arXiv:1711.11383*, 2017.
- [179] S. Goldman and Y. Zhou, “Enhancing supervised learning with unlabeled data,” in *ICML*. Citeseer, 2000, pp. 327–334.
- [180] C. Chen, Y. Liu, M. Kumar, J. Qin, and Y. Ren, “Energy consumption modelling using deep learning embedded semi-supervised learning,” *Computers & Industrial Engineering*, vol. 135, pp. 757–765, 2019.
- [181] S. Baluja, “Probabilistic modeling for face orientation discrimination: Learning from labeled and unlabeled data,” in *Advances in Neural Information Processing Systems*, 1999, pp. 854–860.
- [182] K. P. Nigam, “Using unlabeled data to improve text classification,” Carnegie-Mellon Univ Pittsburgh PA School of Computer Science, Tech. Rep., 2001.
- [183] G. G. Towell, “Using unlabeled data for supervised learning,” in *Advances in neural information processing systems*, 1996, pp. 647–653.
- [184] B. M. Shahshahani and D. A. Landgrebe, “The effect of unlabeled samples in reducing the small sample size problem and mitigating the hughes phenomenon,” *IEEE Transactions on Geoscience and remote sensing*, vol. 32, no. 5, pp. 1087–1095, 1994.
- [185] K.-F. Wong, M. Wu, and W. Li, “Extractive summarization using supervised and semi-supervised learning,” in *Proceedings of the 22nd international conference on computational linguistics (Coling 2008)*, 2008, pp. 985–992.
- [186] A. Blum and T. Mitchell, “Combining labeled and unlabeled data with co-training,” in *Proceedings of the eleventh annual conference on Computational learning theory*, 1998, pp. 92–100.
- [187] R. D. Ball *et al.*, “Parton distributions with LHC data,” *Nucl. Phys. B*, vol. 867, pp. 244–289, 2013.
- [188] J. De Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaitre, A. Mertens, M. Selvaggi, D. . Collaboration *et al.*, “Delphes 3: a modular framework for fast simulation of a generic collider experiment,” *Journal of High Energy Physics*, vol. 2014, no. 2, p. 57, 2014.
- [189] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-kt jet clustering algorithm,” *Journal of High Energy Physics*, vol. 2008, no. 04, p. 063, 2008.

- [190] Cacciari, Matteo and Salam, Gavin P and Soyez, Gregory, “Fastjet user manual,” *The European Physical Journal C*, vol. 72, no. 3, p. 1896, 2012.
- [191] M. Lehmann, “b-tagging algorithms and their performance at ATLAS,” *arXiv preprint arXiv:0809.4896*, 2008.
- [192] D. de Florian *et al.*, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector,” vol. 2/2017, 10 2016.
- [193] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuszi, M. Spira, L. H. C. S. W. Group *et al.*, “Standard model Higgs-boson branching ratios with uncertainties,” *The European Physical Journal C*, vol. 71, no. 9, p. 1753, 2011.
- [194] R. Brun and F. Rademakers, “Root—an object oriented data analysis framework,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 389, no. 1-2, pp. 81–86, 1997.
- [195] J. Therhaag, “TMVA: Toolkit for multivariate data analysis,” *AIP Conf. Proc.*, vol. 1504, no. 1, pp. 1013–1016, 2012.
- [196] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.
- [197] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for activation functions,” *arXiv preprint arXiv:1710.05941*, 2017.
- [198] A. L. Maas, A. Y. Hannun, A. Y. Ng *et al.*, “Rectifier nonlinearities improve neural network acoustic models,” in *Proc. icml*, vol. 30, no. 1. Citeseer, 2013, p. 3.
- [199] D. E. Rumelhart, R. Durbin, R. Golden, and Y. Chauvin, “Backpropagation: theoretical foundations,” *Backpropagation and Connectionist Theory*. Lawrence Erlbaum, 1992.
- [200] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning internal representations by error propagation,” California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985.
- [201] X.-H. Yu, G.-A. Chen, and S.-X. Cheng, “Dynamic learning rate optimization of the backpropagation algorithm,” *IEEE Transactions on Neural Networks*, vol. 6, no. 3, pp. 669–677, 1995.
- [202] R. Callan, *Essence of neural networks*. Prentice Hall PTR, 1998.
- [203] A. Carling, “Back propagation,” *Introducing Neural Networks*, pp. 133–154, 1992.
- [204] J. Freeman and D. Skapura, “Backpropagation neural networks algorithm applications and programming techniques,” 1992.

- [205] M. A. Otair and W. A. Salameh, "Speeding up back-propagation neural networks," in *Proceedings of the 2005 Informing Science and IT Education Joint Conference*, vol. 1, no. 0002, 2005.
- [206] L. He, Y. Bo, and G. Zhao, "Speech-oriented negative emotion recognition," in *2015 34th Chinese Control Conference (CCC)*. IEEE, 2015, pp. 3553–3558.
- [207] J. Ge, J. Sha, and Y. Fang, "An new back propagation algorithm with chaotic learning rate," in *2010 IEEE International Conference on Software Engineering and Service Sciences*. IEEE, 2010, pp. 404–407.
- [208] S. Iranmanesh, M. A. Mahdavi *et al.*, "A differential adaptive learning rate method for back-propagation neural networks," *World Academy of Science, Engineering and Technology*, vol. 50, no. 1, pp. 285–288, 2009.
- [209] Z. Zainuddin, N. Mahat, and Y. A. Hassan, "Improving the convergence of the backpropagation algorithm using local adaptive techniques." in *International Conference on Computational Intelligence*. Citeseer, 2004, pp. 173–176.
- [210] M. S. Al_Duais and F. Mohamad, "A review on enhancements to speed up training of the batch back propagation algorithm," *Indian Journal of Science and Technology*, vol. 9, no. 46, pp. 1–10, 2016.
- [211] M. Pfister and R. Rojas, "Speeding-up backpropagation-a comparison of orthogonal techniques," in *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, vol. 1. IEEE, 1993, pp. 517–523.
- [212] M. Nielsen. Neural networks and deep learning. [Online]. Available: <http://neuralnetworksanddeeplearning.com/chap2.html>
- [213] S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," *Applied artificial intelligence*, vol. 17, no. 5-6, pp. 375–381, 2003.
- [214] J. Huang, Y.-F. Li, and M. Xie, "An empirical analysis of data preprocessing for machine learning-based software cost estimation," *Information and software Technology*, vol. 67, pp. 108–127, 2015.
- [215] J. Keung, E. Kocaguneli, and T. Menzies, "Finding conclusion stability for selecting the best effort predictor in software effort estimation," *Automated Software Engineering*, vol. 20, no. 4, pp. 543–567, 2013.
- [216] L. Angelis and I. Stamelos, "A simulation tool for efficient analogy based cost estimation," *Empirical software engineering*, vol. 5, no. 1, pp. 35–68, 2000.
- [217] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

- [218] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, “scikit-learn: Machine learning in python,” *journal of machine learning research*, vol. 12, p, 2011.
- [219] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, and G. S. Corrado, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, software available from tensorflow.org. [Online]. Available: <https://www.tensorflow.org/>
- [220] R. H. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, “Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit,” *Nature*, vol. 405, no. 6789, pp. 947–951, 2000.
- [221] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, “What is the best multi-stage architecture for object recognition?” in *2009 IEEE 12th international conference on computer vision*. IEEE, 2009, pp. 2146–2153.
- [222] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Icml*, 2010.
- [223] J. Feng and S. Lu, “Performance analysis of various activation functions in artificial neural networks,” in *Journal of physics: conference series*, vol. 1237, no. 2. IOP Publishing, 2019, p. 022030.
- [224] B. Ding, H. Qian, and J. Zhou, “Activation functions and their characteristics in deep neural networks,” in *2018 Chinese control and decision conference (CCDC)*. IEEE, 2018, pp. 1836–1841.
- [225] F. Agostinelli, M. Hoffman, P. Sadowski, and P. Baldi, “Learning activation functions to improve deep neural networks,” *arXiv preprint arXiv:1412.6830*, 2014.
- [226] J. Liang and R. Liu, “Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network,” in *2015 8th international congress on image and signal processing (CISP)*. IEEE, 2015, pp. 697–701.
- [227] R. Caruana, S. Lawrence, and L. Giles, “Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping,” *Advances in neural information processing systems*, pp. 402–408, 2001.
- [228] L. Prechelt, “Early stopping-but when?” in *Neural Networks: Tricks of the trade*. Springer, 1998, pp. 55–69.
- [229] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [230] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” *arXiv preprint arXiv:1207.0580*, 2012.

- [231] Y. Bengio, “Practical recommendations for gradient-based training of deep architectures,” in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 437–478.
- [232] W. Forst and D. Hoffmann, *Optimization—theory and practice*. Springer Science & Business Media, 2010.
- [233] I. Goodfellow, Y. Bengio, and A. Courville, “Deep learning (adaptive computation and machine learning series),” *Cambridge Massachusetts*, pp. 321–359, 2017.
- [234] L. N. Smith, “Cyclical learning rates for training neural networks,” in *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2017, pp. 464–472.
- [235] S. Wen, S. Xiao, Y. Yang, Z. Yan, Z. Zeng, and T. Huang, “Adjusting learning rate of memristor-based multilayer neural networks via fuzzy method,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 38, no. 6, pp. 1084–1094, 2018.
- [236] C. W. Murphy, “Class imbalance techniques for high energy physics,” *SciPost Physics*, vol. 7, no. 6, p. 076, 2019.
- [237] S. Choi, M. Muhlleitner, and P. Zerwas, “Theoretical basis of Higgs-spin analysis in $H \rightarrow \gamma\gamma$ and $Z\gamma$ decays,” *Physics Letters B*, vol. 718, no. 3, pp. 1031–1035, 2013.