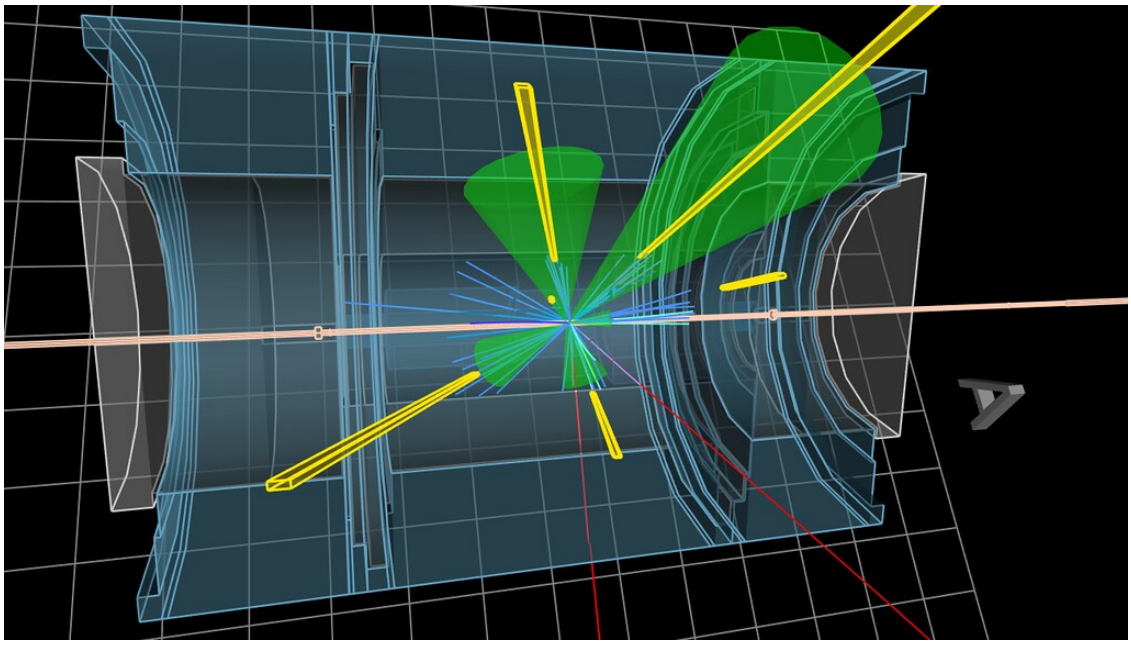
CHALMERS
UNIVERSITY OF TECHNOLOGY



Testing the Bell inequality for Higgs boson decay

With supervised learning Multi-Variate Analysis classifiers

Master's thesis in Physics

Axel Pantzare

Department of Physics

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023
www.chalmers.se

Degree project report 2023

Testing the Bell inequality for Higgs boson decay

With supervised learning Multi-Variate Analysis classifiers

Axel Pantzare

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Physics
Chalmers University of Technology
Gothenburg, Sweden 2023

Testing the Bell inequality for Higgs boson decay
With supervised learning Multi-Variate Analysis classifiers
Axel Pantzare

Supervisor: Mats Granath, Department of Physics
Examiner: Mats Granath, Department of Physics

Master's thesis in Physics 2023
Department of Physics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 31 772 1000

Cover: Visualisation of event at the ATLAS detector.

Typeset in LaTeX
Gothenburg, Sweden 2023

Testing the Bell inequality for Higgs boson decay
With supervised learning Multi-Variate Analysis classifiers
AXEL PANTZARE
Department of Physics
Chalmers University of Technology

Abstract

The Higgs boson was discovered in the ATLAS and CMS detectors at LHC in 2012. Since then, more and more have been learned about its properties including its mass. Bell introduced a test to verify if particles behaved quantum mechanically or classically. This was later modified to be able to test a three-level system such as the Higgs boson decay to W bosons. The data from ATLAS open data was used to train neural networks to classify these events and filter out the background on simulated data and then use these to filter out the actual data. This was then tested in the generalized Bell inequality for a three-level system. It was found that the actual data filtered out by the network broke the inequality with over 4σ but was far away from the expected value. This can be due to precision lower than for the training and validation data because of differences between simulations and actual data but also because the assumptions of the Higgs boson being at rest in the lab frame were insufficient.

Keywords: Bell inequality, Supervised learning, Higgs boson decay, CGLMP inequality.

Acknowledgements

First, I want to thank Hok-Chuen "Tom" Cheng at The Chinese University of Hong Kong that introduced me to the subject in the spring of 2022 and gave feedback on my first draft. I also want to thank my supervisor Mats Granath for his advice and the computational time on both the Vera and Alvis clusters. Computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing (SNIC) at Chalmers Centre for Computational Science and Engineering (C3SE), partially funded by the Swedish Research Council through grant agreements no. 2022-06725 and no. 2018-05973.

Axel Pantzare, Gothenburg, May 2023

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

ATLAS	A Toroidal LHC ApparatuS
CERN	Conseil Européen pour la Recherche Nucléaire
CHSH	Clauser-Horne-ShimonyHol
CGLMP	Collins-Gisin-Linden-MassarPopescu
DNN	Deep Neural Network
JVT	Jet Vertex Tagging
LHC	Large Hadron Collider
MC	Monte Carlo
MVA	Multi Variate Analysis
PCA	Principal component analysis
SSP	Spontanious Symmetry Breaking
t-SNE	t-Distributed Stochastic Neighbor Embedding
QFT	Quantum Field Theory

Nomenclature

Below is the nomenclature of indices, particles, parameters, variables and groups that have been used throughout this thesis.

Indices

i	Index for the leptons
j	Index for the jets
k	Index for the different coordinate axis
l	Index for the different layers.
m	Index for the different neurons.

Particles

l	Leptons
ν_l	Neutrino of flavor l
μ	Muon
e	Electron
b	Bottom quark
W	W boson
g	Gloun
τ	Tauon
c	Charm quark
Z	Z boson
H	Higgs boson

Parameters

α	Learning rate
Max η	Maximum pseudorapidity of jet finder
r	Cone radius of jet
Min p_T	Minimum transverse momentum for jet finder
nη	Range of jet finder
nϕ	Granularity of jet finder

Variables

L	Weak isospinn
Y	Weak hypercharge
Φ	Scalar field
B	B Boson
ν	Constant for Higgs potential
λ	Constant for Higgs potential
v	Vacuum expectation value of Higgs potential
η	Pseaudorapidity
θ	Angle between the particle three momenta and the beam
S	Value of bell inequality
a	Detector settings at side a
b	Detector settings at side b
ξ	Cosine of azimuth angle of lepton.
I	CGLMP Inequality value or contribution from term k
σ	Standard deviation of data
$\tilde{\sigma}$	Standard deviation of mean
n	Number of events selected
N	Size of dataset
y	Output from neuron
x	Input to neuron
w	Weight to input
b	Bias for neuron
$f(y)$	Activation function
$p(y)$	Predicted value
μ	Mean of data

m_{LL}	Combined mass of leptons
$goodjet$	Number of "goodjets"
MET	Missing Transverse energy
$ptLL$	Transverse momentum of
$goodbjet$	Number of B jets
FP	False positives
TP	True positives
FN	False negatives
TN	True negatives
P	Precision
BA	Binary accuracy

Groups

$U(1)$	Unitary group of dimension 1. This corresponds to rotations on the unit circle.
$SU(2)$	Special Unitary group of dimension 2. The group is represented by 2 by 2 unitary matrices with determinant 1.
$SU(3)$	Special Unitary group of dimension 3. The group is represented by 3 by 3 unitary matrices with determinant 1.

Contents

List of Figures

List of Tables

1
Introduction

The Higgs boson was proposed in 1964 to explain how all other particles in the standard model acquired their mass [5]. This was later discovered in 2012 by the ATLAS and CMS detectors at LHC in Cern [6]. What laws of physics affect this particle's decay has been a question no one has been able to answer. The answer to this question of the Higgs boson decay would test the theories of quantum mechanics at higher energies than probed before.

1.1 Background

One of the natural consequences of quantum mechanics is the property of entanglement. This means that two particles that are apart from each other and in a superposition of states can "feel" when the other particle is being measured and the superposition collapses into a single state. This has long been known to violate the inequalities formed by Bell [7]. This follows naturally from quantum physics but goes against the realist approach and classical physics where the two particles should be isolated from each other:

These theories have been tested in multiple different experiments for different two-level systems and it has been proven to violate the Bell inequality in for example superconduction, ions, photons, and nitrogen-vacancy centers. In addition, also some three-level systems using photons [8]. In this thesis, the aim is to test the Bell inequality for Higgs boson decay to two W bosons which will be at the highest energy this far.

Last year the Bell inequality was tested for the decay from Higgs to two W bosons for simulated data. It was concluded that this data pointed towards that this decay violates the inequality which means that it decays quantum mechanically, this is expected but has never been tested at these energies in real experiments before [8]. In the article, the author proposes that this should be possible to investigate with the data already available from the Large Hadron Collider (LHC). Earlier this year it was also shown analytically with quantum corrections that this should be the case [9].

1.2 Purpose and goals

This thesis aims to use experimental data to decide if Higgs boson decay to WW bosons breaks the Bell inequality or not with a reasonable confidence level. The decay mode to WW bosons is chosen since the decay products are heavier than for example the mode to two bottoms or gluon, which might mean that it requires higher energies. The WW mode was chosen over the ZZ mode since the branching ratio is much higher for WW which will give more data to analyze.

1.3 Limitations

The scope of this thesis is only 30 ECTS credits which roughly corresponds to working 20 weeks full-time. The limitation in terms of data is that it is only the 13 TeV open data set that is accessible. This is relatively small and might give large uncertainties of the calculated numbers.

1.4 Problem formulation

Does the Higgs boson decay into two W bosons break the Bell inequality? How can one filter out the H to WW events from the background?

1.5 Ethics

During this project, ethics have been taken into account by crediting the people and projects this thesis builds on. I have also been clear about what drawbacks this research has to not mislead anyone. Furthermore, the code is also uploaded to GitHub for others to see and continue building on. Since this research is not directly applicable to any industry or in other way human-facing no extra caution has been taken in the research design in terms of ethics.

2
Theory

In this chapter, first, the particle physics involved is introduced in section 2.1 and especially discusses the Standard Model, The Higgs boson, how it is formed and decays, and the Bell inequality in section 2.1.5. Furthermore, the ATLAS detector is explained in section 2.2. Lastly, the data in is presented section 2.3, and the theory behind the multi-variate analysis (MVA) classifiers is presented in section 2.4.

2.1 Particle physics

All things we interact with are built up of atoms. These atoms are in turn built up by protons, neutrons, and electrons. The protons and neutrons are built up by quarks. Electrons and quarks are examples of elementary particles. It is the elementary particles that build up most matter we know. The theoretical model that describes these fundamental particles and their properties is called the Standard Model. To validate and test this model multiple particle colliders have been built and the largest one which generates the highest energies in collisions is the Large Hadron Collider (LHC) in Cern [10].

2.1.1 Standard model

The Standard Model builds on experimental observations and mathematical derivations for which particles are needed to build up our world. The Standard Model builds on quantum field theory (QFT) and gauge invariance to explain all of these particles and also the forces (except gravity) that makes them interact. The particles can be divided into two classes, half-integer spin particles known as fermions, and integer spin particles known as bosons. These play different roles where the fermions build up all matter and the bosons are the force-carrying particles that give rise to the fundamental forces.

QFT is the theoretical study of particle physics which combine quantum mechanics, special relativity, and classical field theory. This theory treats particles as an excitation of their respective field.

The Standard Model is based on the symmetry group $SU(3) \times SU(2) \times U(1)$, where $SU(3)$ describes the strong force between quarks, $SU(2)$ describes the electroweak force, and $U(1)$ describes the electromagnetic force [11]. Each of these groups is associated with a set of gauge fields or gauge bosons, which mediate the interactions

between particles.

The most relevant particles for this thesis are the electron and muon neutrino, the electron, the muon, and the W and H bosons.

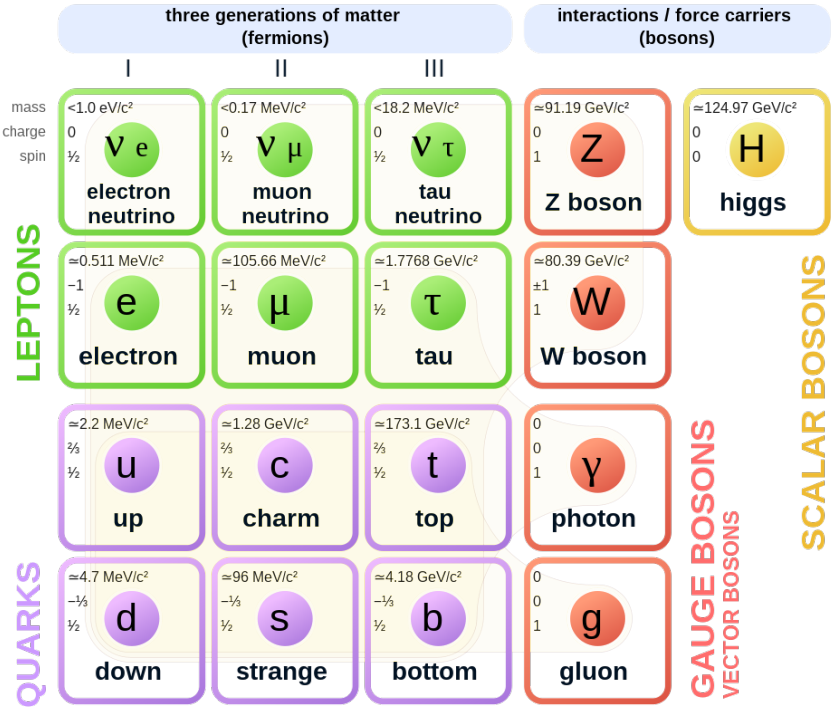


Figure 2.1: The Standard Model as a visual representation with the 12 fermions and 5 bosons. Note the spin, mass, and charges of the H, W, e, and μ. Creative Commons Attribution 3.0 Unported license [1].

2.1.2 Higgs boson

In the beginning, all particles were massless, but a fraction after the big bang some acquired mass. This troubled physicists for a long time before Peter Higgs proposed that spontaneous symmetry breaking (SSP) and a new quantum field would solve this problem [5]. SSP can be seen as that the universe was initially in a symmetric unstable state but was then spontaneously moved into an asymmetrical state which made some but not all particles acquire mass while the equations were unchanged. The new quantum field was what we now know as the Higgs field and explains why some but not all particles have mass.

The Higgs field is a scalar field that is coupled to the gauge bosons of the electroweak force. The gauge symmetry group associated with the electroweak force is $SU(2)_L \times U(1)_Y$, where $SU(2)_L$ describes the weak isospin symmetry and $U(1)_Y$ describes the weak hypercharge symmetry. A visual representation of the group $SU(2)$ is rotations on the surface of a 3D sphere and for $U(1)$ the visual representation is rotations on the unit circle. The electroweak gauge bosons consist of three $SU(2)_L$ gauge bosons, W^+, W^-, Z^0.

The Higgs field is a doublet of $SU(2)_L$ and has weak hypercharge $Y = 1$, and can be written as:

$$\Phi = \begin{pmatrix} \Phi^+ & \Phi^0 \end{pmatrix} \tag{2.1}$$

where Φ^+ and Φ^0 are complex scalar fields.
The Higgs potential is the potential energy in the Higgs field and is given by:

$$V(\Phi) = \nu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{2.2}$$

where $\nu^2 < 0$ and $\lambda > 0$ are constants that determine the shape of the potential [12], ν^2 having an inpact of the center of the potential and λ impacting the edges. The Higgs potential has a "Mexican hat" shape, with the minimum of the potential occurring at a non-zero value of the Higgs field, breaking the electroweak symmetry and giving mass to the electroweak gauge bosons and fermions. The minimum of the Higgs potential occurs when the potential fulfills:

$$|\Phi| = \frac{v}{\sqrt{2}} = \sqrt{-\frac{\nu^2}{2\lambda}} \tag{2.3}$$

where v is the vacuum expectation value (VEV) of the Higgs field, a non-zero value that determines the masses of the gauge bosons and fermions [13].

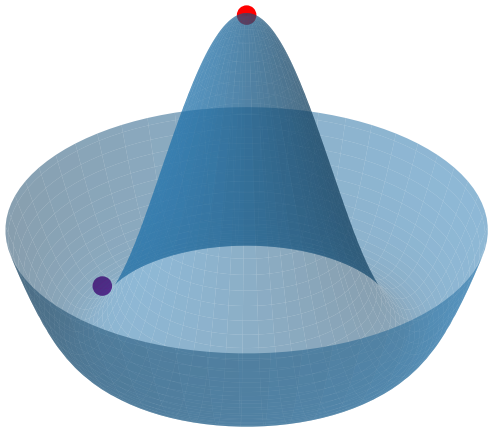


Figure 2.2: The Higgs potential with a red dot at the local maxima and a purple dot at the local minima. Since the local minima is rotationally symmetric this shows the concept of spontaneous symmetry breaking.

Spontaneous symmetry breaking can be seen when the ball moves from the top of the potential (in a symmetric unstable state), red ball to the valley of the potential (stable asymmetric state), purple ball, and thereby breaks the symmetry. With this shift in the Higgs field the gauge bosons gain mass. The field that gives the potential is called the Higgs field and is therefore needed for the Higgs mechanism. Since an excitation of a field is regarded as a particle this gives rise to the Higgs boson.

In summary, the Higgs mechanism is a key component of the Standard Model and is responsible for giving mass to the particles of the electroweak force. The math-

ematical framework behind the Higgs mechanism relies on the symmetry group $SU(2)_L \times U(1)_Y$, the Higgs field, and the Higgs potential. In summary, the Higgs field is a scalar field that is coupled to the gauge bosons of the electroweak force, and the Higgs potential has a "Mexican hat" shape with a non-zero vacuum expectation value that breaks the electroweak symmetry and gives mass to the particles.

2.1.3 Higgs Decay

The Higgs boson can decay in multiple ways. The most common is two bottom quarks followed by two W bosons. Some less common is to two gluons, touons, charms, and Z bosons, all major decay modes are displayed in table 2.1.

Table 2.1: Branching ratios for the Higgs bosons decay into lighter products [3]. There are also a few other decay routes that are neglected here since their contributions are minor.

Decay products	Branching ratio [%]
$b\bar{b}$	57
WW	22
gg	9
$\tau\tau$	6
$c\bar{c}$	3
ZZ	3

Since the W boson is about 20 times heavier than the bottom quark it means that it should couple stronger to the Higgs field and thereby make it easier to investigate the Higgs boson. The downside might be the limitation in statistics since it should be about a third as many events of that kind. But since it is about 10 more common than the Z boson decay it should be the best option to evaluate the Higgs bosons properties.

When the W^+ boson decays it sends out a lepton in the direction of its spin and when the W^- boson decays it sends out a lepton in the opposite direction of its spin [14]. This means that the directions of the leptons in the final state should indicate the spin of the W boson and therefore it should be possible to detect if the W bosons were entangled or not.

Since one H is lighter than two W, one of the W has to be off-shell which means that this breaks the energy-momentum conservation for a short period of time and therefore can be viewed as a virtual particle. The full decay from the Higgs to the detected products is $H \longrightarrow WW \longrightarrow e + \mu + \nu_e + \nu_\mu$. We only want to use the decay into e and μ since $H \rightarrow \tau\tau$ has a significant contribution without decaying into WW first and we want them to have different type to eliminate contributions from pair production. Since the ATLAS detector can not detect neutrinos are they ignored and replaced by a "missing energy" term. If we symbolize the off-shell boson with

an $*$ the decay can be written as

$$\mathrm{H} \longrightarrow \mathrm{WW}^* \longrightarrow \mathrm{e}^{\pm} + \mu^{\mp} + \mathrm{ME}. \tag{2.4}$$

Since the Higgs is a zero spin particle and the W has ± 1 and charge they have to have opposite charges which gives the leptons opposite charges.

2.1.4 Jets

When the two protons collide a hot and dense medium is formed, which is called a quark-gluon plasma. This is characterized by emerging quarks and gluons. Sometimes these fragment into highly collimated particle streams called jets. These jets carry some of the energy and momentum from the collision so lighter particles can form.

The jets can be classified in different ways, for example, jets over 60 GeV need to be restricted. Too many of them carries away too much energy, which makes it impossible to form a Higgs boson, therefore this kind of jet is called a good jet in this report. Jets have bad directional accuracy which makes it hard to determine from which collision vertex a single jet originates. To solve this, Jet Vertex Tagging (JVT) is used to determine how much energy from one track that is from every vertex [15].

One special type of jet that is called a b-jet is extra interesting since it is generally more energetic than other jets. The difference between a b-jet and other jets is that it contains one or more bottom quarks. To classify this ATLAS uses an algorithm called MV2c10 which uses multi-variate analysis to calculate the likelihood of a jet containing a bottom quark. This algorithm has shown great accuracy which has made it the main algorithm for b-tagging at ATLAS. In this report, a value over 0.17 is considered a b-jet and means that the threshold is relatively low which is important in this analysis since too many b-jets would mean we can not get any Higgs in the decay [16].

2.1.5 Bell inequality and CHSH

When the field of quantum mechanics started to gain traction a kind of test was needed to prove that systems behaved quantum mechanically. One such test became the Bell test to prove non-locality or as it is more commonly known now: entanglement [7]. Entanglement is a phenomenon that means that measuring one particle instantaneously (independent of distance) affects another particle. This would not be allowed in classical theories since nothing can travel faster than the speed of light. This means that the state of two entangled particles can not be described individually but have to be described together. The usual formulation of the Bell inequality for a two-level system is called the CHSH inequality and reads [17],

$$S = E(a,b) - E(a,b') + E(a',b) + E(a',b') \leq 2, \tag{2.5}$$

where E is the quantum correlation expectation value of measurement outcomes of different particles with different states. Detector a and b are different sides of

the same experiment with the primed coordinates being different detector settings. These detector settings can, for example, be the polarization of two entangled photons [17].

For a fully entangled state, this inequality should take the value $\sqrt{8} \approx 2.8$ which exceeds the limit 2 set by classical theories which means that quantum mechanical systems break the CHSH inequality [18].

2.1.6 3 level system and CGLMP inequality

Since the Higgs boson has zero spin and is decaying to two W bosons that can have integer spin between -1 and 1 this can be described as a 3-level system. The system can therefore be described as

$$|\psi_s\rangle = \frac{1}{\sqrt{3}}(|-\rangle\,|+\rangle + |0\rangle\,|0\rangle + |+\rangle\,|-\rangle). \tag{2.6}$$

The W bosons will then decay into a lepton and a neutrino where these leptons will be emitted in the direction of the spin for the W^+ and in the opposite direction of the spin for the W^- to maximally violate chirality since this is a weak decay [14]. This makes it possible to measure the direction of the spin by measuring the lepton's directions. The leptons $l^{\pm} \in \{e^{\pm}, \mu^{\pm}\}$ since we are excluding the τ to get higher sensitivity of the search [6]. The leptons will be of different charges since the Higgs is neutral.

The optimal test of the Bell inequality for a three-level system is the CGLMP inequality [8] which is a generalization of the CHSH inequality which expectation value can be computed by

$$I = \frac{3}{\sqrt{8}}\langle\xi_x^+\xi_x^- + \xi_y^+\xi_y^-\rangle + 25\langle\Big((\xi_x^+)^2 - (\xi_y^+)^2\Big)\Big((\xi_x^-)^2 - (\xi_y^-)^2\Big)\rangle + 100\langle\xi_x^+\xi_y^+\xi_x^-\xi_y^-\rangle \tag{2.7}$$

where ξ_k^n is the cosine for the angle of the lepton with charge n with respect to the axis k [8] in the rest frame of the Higgs. However, since the Higgs is about 100 times heavier than the protons collided most of the energy will go into mass and it can therefore be assumed that it is nearly at rest in the lab frame [19]. The only axis used is x and y since that is the only direction that has full resolution because of how the ATLAS detector is built. By plotting the CGLMP value for 5000 different pairs of random angles we get a general idea of the dependence on the angle difference. We see that if the angle difference is small or close to π we get high values of the inequality and if there is a more open angle between the leptons we get a lower value.

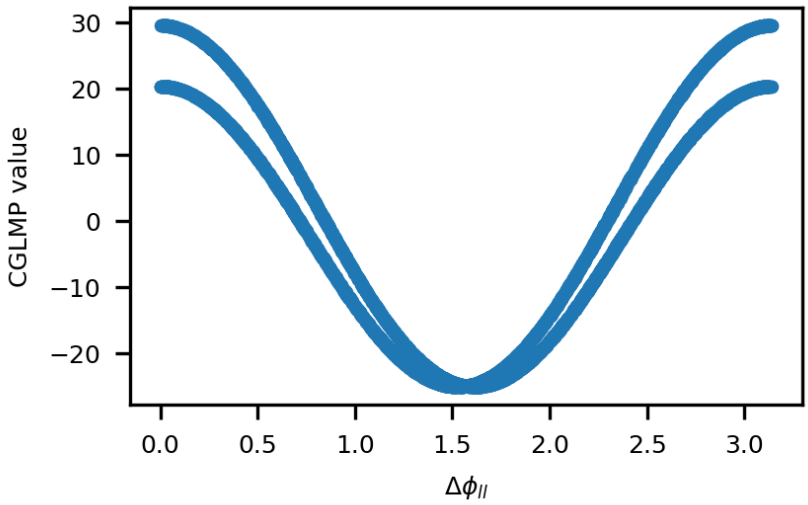


Figure 2.3: CGLMP value as a function of $\Delta\phi$. Smaller angle differences between the leptons give a larger value while a more open angle gives a smaller value. Since the quantity depends on both angles and not just the difference it is two lines instead of one.

The statistical uncertainty can be calculated as the sum of the variances for the three different averages and the sum of their covariance.

$$\begin{aligned}\sigma^2 = \left(\frac{3}{\sqrt{8}}\right)^2 \mathrm{var}(I_1) + 25^2\mathrm{var}(I_2) + 100^2\mathrm{var}(I_3) + \\ \frac{3}{\sqrt{8}} \cdot 25\mathrm{cov}(I_1, I_2) + \frac{3}{\sqrt{8}} \cdot 100\mathrm{cov}(I_1, I_3) + 25 \cdot 100\mathrm{cov}(I_2, I_3)\end{aligned} \tag{2.8}$$

Since these are all mean values the error of means will be used which means that we divide by the square root of the number of samples one more time.

$$\tilde{\sigma} = \frac{\sigma}{\sqrt{n}} \tag{2.9}$$

The maximum value for this quantity for a classical system should be 2 and for a fully entangled system should be ≈ 2.9 [8].

2.2 ATLAS detector

The ATLAS detector is the largest detector at LHC and includes of six different detection subsystems in multiple layers with calorimeters that detect the energy of particles. By applying a magnetic field, the charge, and mass can be calculated by measuring how they deposit energy over the detectors calorimeters.

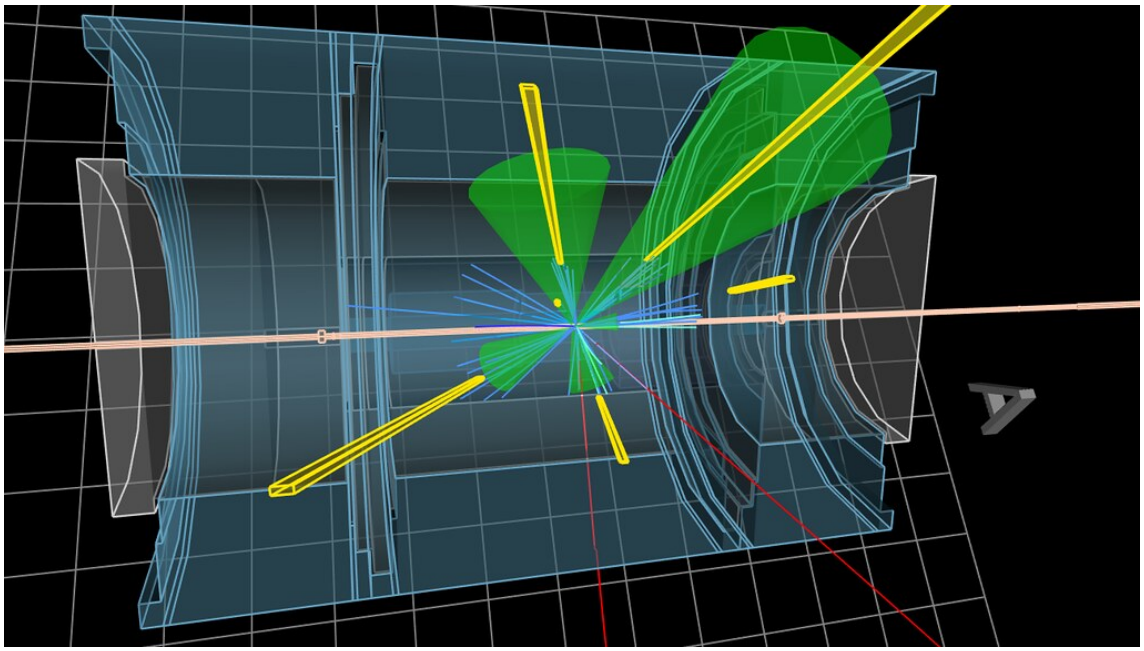


Figure 2.4: Visualisation of the reconstruction of jets and particles in ATLAS. The yellow beams are the jets, the green cones are the area around the jets that take a group of particles and combine into one jet and the blue rays are each individual particle. Creative Commons NonCommercial-ShareAlike 2.0 license [2].

The coordinate system at ATLAS has the x-axis pointing towards the middle of the beam ring, the y-axis pointing up from the ground and the z-axis pointing along the beam direction to form a right-handed system. Instead of measuring the angle θ that the decay product deflects from the beam the pseudo rapidity η is used which is calculated through

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{2.10}$$

2.3 Data

The data is composed of 49 different data channels such as the lepton's/jet's momentum and directions. A full list of the data channels can be viewed in table 2.2.

Table 2.2: Variables used for the classification data along with descriptions for what they mean [4]. They are grouped together in three groups, attributes that describe the whole event, lepton attributes for both leptons and jet attributes for the three most energetic jets.

Chanel name	Chanel description
m_{ll}	Invariant or total mass of the final products
pt_{ll}	Total transverse momentum of leptons
$\Delta\phi_{ll}$	Angle difference in the xy-plane between leptons
$\Delta\phi_{ll}MET$	Angle difference in the xy-plane between the lepton center of mass and the missing energy
MET	Missing energy
m_T	Dilepton transverse mass
jet_n	Number of jets
$goodjet_n$	Number of jets with pt over 60 GeV and η larger than 2.4
$goodbjet_n$	Number of jets with pt over 60 GeV, η between 2.4 and 2.5 and MV2C10 value over 0.176 (85%)
pt_i	transverse momentum for lepton i
η_i	pseudorapidity momentum for lepton i
E_i	Energy for lepton i
ϕ_i	angle in the xy-plane for lepton i
lep_charge_i	Charge for lepton i
lep_type_i	Type for lepton i ($\mu = 13$, $e = 11$)
$lep\_trackd0pvunbiased_i$	d_0 of track associated with lepton at point of closest approach
$lep\_tracksigd0pvunbiased_i$	d_0 Significance of the track associated at the point of closest approach
lep_z0_i	z-coordinate of the track associated with the lepton wrt. primary vertex
$lep\_ptcone30_i$	Scalar sum of track pt in a cone of $R = 0.3$ around lepton used for tracking isolation
$lep\_etcone20_i$	Scalar sum of track Et in a cone of $R = 0.2$ around lepton used for calorimeter isolation
$jetpt_j$	Transverse momentum of jet
$jet\eta_j$	Pseudorapidity of jet
$jetE_j$	Energy of jet
$jet\phi_j$	Angle of the jet in the xy-plane
jet_MV2c10_j	Output from the multivariate b-tagging algorithm of the jet
jet_jvt_j	Jet vertex discriminant of jet

The training data for this project was a combination of the Monte Carlo (MC) data available as open source through ATLAS Open Data and additional data that had been simulated with Pythia [20] to closely resemble the actual data. It is building on the Standard Model with next-to-leading order (NLO) corrections. The jets have been simulated with Fastjet [21] with the standard settings. This is the same software that was used for the same type of events as in the open data.

The data set from the collider has the same data channels as the MC data but these have been reconstructed from the detector data. The data is taken during run 2 which lasted between 2015 to 2018 and the Center of Mass (CM) energy was 13 TeV [22].

2.4 MVA Classifier

Since the signal (H→WW) and the background are very similar an MVA classifier is needed to select the parts that are most likely to be signal. There are multiple different options for MVA classifiers but the ones discussed here are Dense Neural Networks (DNN), Principal Component Analysis (PCA), and 1D convolutional networks. There are indefinitely many ways of using these but some general background is presented.

2.4.1 PCA

Since some of the features are correlated this will give rise to multicollinearity. To avoid this PCA has been used. PCA means that an N-dimensional hyperplane is fitted to the data to explain as much of the variance as possible. This will reduce the number of features and decrease the correlation between features [23].

2.4.2 DNN

A deep neural network is a machine learning process that is, simply put, made to learn in a similar way as the brain. It uses layers of neurons that transform the input data to closely resemble the output data of the training examples [24].

2.4.2.1 Single neuron

A single neuron is the smallest element that the neural network is built up of. It takes in the output of the previous layer (or the features in the first) and transforms it with a linear function applying weights w_m to all input data and a bias b_m such as

$$y_m^l = \sum w_m x_m + b_m, \tag{2.11}$$

for neuron m in layer l. An non-linear activation function is then used on y_m^l to get the output [24].

2.4.2.2 Activation function

To make sure the output after each neuron signals either on or off or a probability of it the activation function is used to map y_m^l to [0,1]. There are several different but the most common ones are “ReLu”, “Softmax” and “Sigmoid” [25]. The sigmoid is defined as

$$f(y) = \frac{1}{1 + e^{-y}} \tag{2.12}$$

ReLu is a combination of a linear activation and a binary activation [25]. It stays at zero until the input is 0 and then increases linearly

$$f(y) = \begin{cases} 0 & y \leq 0 \\ y & 0 \leq y. \end{cases}$$

The upside of this is that all neurons are not activated at the same time which makes it suitable for use in hidden layers. The softmax activation function is a mathematical function that takes a vector of real-valued inputs and produces a vector of values that represent the probabilities of each class. It exponentiates the input values and normalizes them in such a way that they range from 0 to 1 and sum up to 1, effectively transforming them into probabilities. This transformation enables the interpretation of the output as a probability distribution over multiple classes, facilitating the assignment of a single class label among the available options. This makes it especially good for multiple-class classification whereas a single sigmoid works better for binary classification [25].

2.4.2.3 Structure

The DNN is built up of multiple layers with a dropout between each layer which sets a portion of the data to 0 and then re-normalizes the sum. This prevents the model from overfitting. The last layer has only one neuron to project the data to one number which is compared to the label to evaluate if the model did a successful classification or not [26].

2.4.3 1D convolutional network

A 1D convolutional network exhibits both similarities and differences when compared to a traditional DNN. While both architectures are used for learning patterns in data, a 1D convolutional network introduces unique characteristics that make it particularly suitable for processing sequential data [26].

One notable distinction lies in the connectivity pattern between layers. In a DNN, every node in a layer is connected to every node in the subsequent layer, forming a fully connected structure. In contrast, a 1D convolutional network employs local connectivity, where each node in a layer is only connected to a subset of nodes in the next layer [26]. This design mimics the way information flows in the human brain and allows the network to capture local patterns and interactions effectively. By reducing the number of connections, 1D convolutional networks can be more computationally efficient, making them advantageous in scenarios with large input sizes or limited computational resources [26].

Furthermore, the organization of data in a 1D convolutional network differs from a DNN. In the context of time series data, the input is typically represented as a sequence or series of values along a single dimension. This temporal aspect is captured by organizing the data into different vectors or channels, where each channel represents a distinct aspect or feature of the time series. For example, in sensor data analysis, different channels can represent various sensor measurements. By scanning over these multiple channels, the 1D convolutional network can detect correlations, similarities, and differences across different aspects of the time series, enabling it to extract meaningful temporal features and patterns. [26].

2.4.4 Loss function

The purpose of the loss function is to tell the network how far away from the right answer the output is. There are multiple different methods for this but in the case of binary classification as this, the most relevant one is binary cross-entropy. This builds on calculating how far away every predicted label is from the actual one and uses the same equation as entropy in statistical mechanics which can be written as

$$BCE = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(p(y_i)) + (1 - y_i)\log(1 - p(y_i)) \tag{2.13}$$

where y_i is the real lable and $p(y_i)$ is the probability that y_i is signal [27].

2.4.5 Optimization

To optimize the network multiple different optimizers can be used. The most common one is adaptive moment estimation (Adam) which is an extension of stochastic gradient descent. This calculates the first and second moments of the gradient to adapt the learning rate (step size) for each weight [28].

3

Methods

The method used to investigate the CGLMP inequality is built on first filtering out as large a proportion of the $H \rightarrow WW$ events as possible and then making the calculation of the CGLMP inequality. To do this the theoretical selection criteria described in section 2.1.3 were first used together with criteria to ensure the quality of the data. After this, two MVA classifiers was trained on the simulated data and then tuned by the expected number of events and used on the real collider data from Cern. The data used was the 13 TeV ATLAS open dataset produced at LHC together with extra simulated data from Pythia [20] and Fastjet [21]. The ALTAS open dataset includes both simulation and experimental data [29].

3.1 Approximation of number of Higgs events

The events that give the main contributions to the data are from five different categories. By fitting the simulated spectra of these different categories to the spectra of the measured data the number of Higgs events in the data can be calculated with the MC weights and detector fractions. With the criteria described below this came out as 425 Higgs to WW events out of the total 21000 events. An example of one spectrum used to approximate the number of Higgs events can be seen in fig. 3.1.

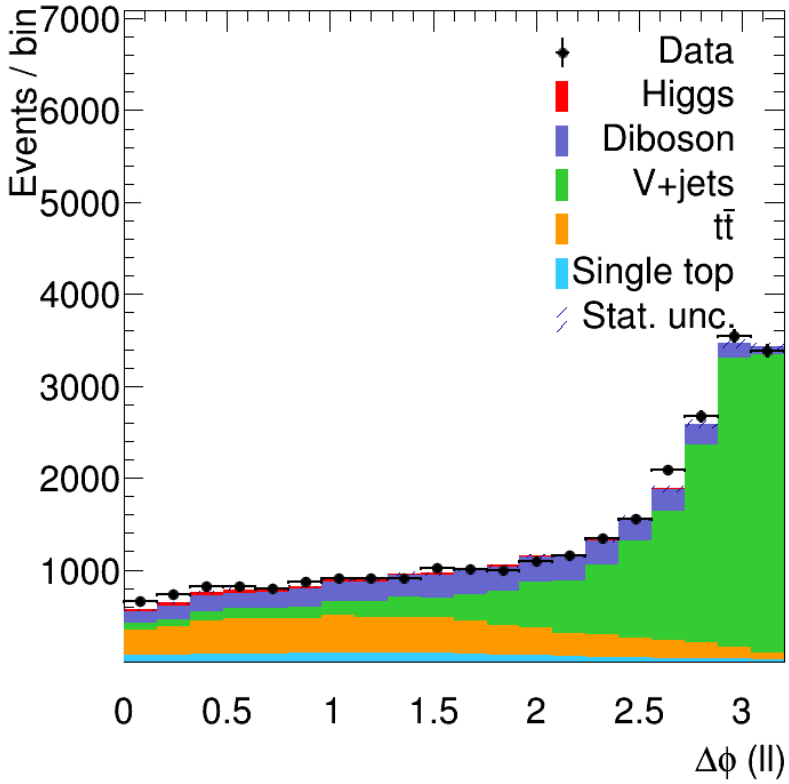


Figure 3.1: The distribution of data for $\Delta\phi(ll)$. Both data and the five different sub-processes are visible. The distribution from the sub-processes is from the simulation and then fitted to the actual data.

3.2 Simulation

The training data is comprised of two sets of simulation data. The first is the ATLAS 13 TeV open data simulated with Pythia at Cern [30]. The second was simulated with Pythia and Fastjet for this thesis work. This was only H→WW since the dataset was quite unbalanced. The settings used were only one production mode which was gluon-gluon fusion to Higgs with CM energy 13 TeV. The only decay mode activated for the Higgs was two W bosons, for the W bosons the decay modes that were activated were only to electron/positron and their respective neutrino and muon/anti-muon with their respective neutrino. For the jets, the standard settings of Max $\eta = 4$, $r = 0.7$, Min $p_T = 10$, excluding neutrinos and other particles that can not be detected, and the range and granularity of the jet finder were set to n$\eta = 80$ and n$\phi = 64$ were used. Before merging the two sets it was checked that all the features looked similar.

3.3 Preprocessing

To filter out a high concentration of H→ WW a multi-step approach was used. First, all the data was filtered with the theoretical selection criteria described in

section 2.1.3 and with criteria to ensure the quality of the data as well as some criteria to filter out the worst background using for example the number of b-jets. Then some extra manipulations of the orientation of the coordinate system were added to the training data to ensure the generativity of the MVA. Lastly, the data were transformed to be of a similar structure for all features.

3.3.1 Selection criteria

As the first filtering, a cut-based selection was made. This consisted of theoretical constraints such as only using events with exactly two differently flavored leptons with opposite charges.

We also want to be sure that it is isolated leptons we are looking at and we can therefore add the criteria that each lepton needs to have transverse momentum over 15 GeV and $ptcone30/pt$ under 0.1 and $etcone20/pt$ under 0.1 and η under 2.5, this means that the particles have some space around it without any other particle which means that we are sure that we can track each lepton individually [4]. Lastly, selection criteria to filter out most of the background were added

$$\begin{aligned} 85 \text{ GeV} > m_{LL} > 5 \text{ GeV} \\ goodjet \leq 1 \\ MET > 10 \text{ GeV} \\ ptLL > 10 \text{ GeV} \\ goodbjet < 2. \end{aligned} \tag{3.1}$$

This is significantly looser than the standard criteria employed for $H \rightarrow WW$ search at cern and therefore letting through more background and signal than before [29].

This left 1.4 million MC events for training events where 760 000 were Higgs.

3.3.2 Extra data manipulations

Since the detector and experiment are symmetric around the beam in the xy-plane the data should be symmetric in the same way. Therefore a random rotation was made to the data and then added to the rest of the training data. The z-axis was also flipped in this example of the data. This should lead to a more general MVA classifier since this should make the choice of the coordinate system more arbitrary.

3.3.3 Transformation

As the last part, a neural network was trained to filter out even more of the background signals. The network was trained on simulation data which have been simulated with detector effects as well which means it should be fairly similar to the experimental data. Before the data is fed into the neural network it is first rescaled with

$$\frac{x - \mu}{\sigma} \tag{3.2}$$

where μ is the mean of the variable, σ is the standard deviation and x is an array of the variable.

The CGLMP inequality will thereafter be calculated with eq. (2.7) for the filtered events to see if it violates the classical limit.

3.4 MVA

To make a classifier that is as accurate as possible multiple methods were tested. Both a DNN using PCA data as well as a 1D convolutional network were trained and tested individually and together.

3.4.1 Development of MVA

First, the number of neurons and layers was tested to find at what number the model could overfit the data to 100% after 200 epochs. When this level was found a dropout was added between each layer and slowly increased until the validation and training scores behaved similarly indicating no overfit.

Different activation functions were also tested, the activation function that gave the highest accuracy was ReLu for the hidden layers and Sigmoid for the output layer.

Binary cross-entropy, calculated as eq. (2.13) was used as a loss function for both networks together with Adam optimizer and learning rate $\alpha = 0.00005$. This value is a bit lower than the standard which seemed a bit high since the network's training scores fluctuated heavily. Only the step with the lowest validation loss was saved and later used.

3.4.2 PCA

To decrease multicollinearity and reduce the number of features as far as possible a PCA was used. However, this seemed to only be able to reduce the data from 49 to 45 features which makes the training of the network slightly easier.

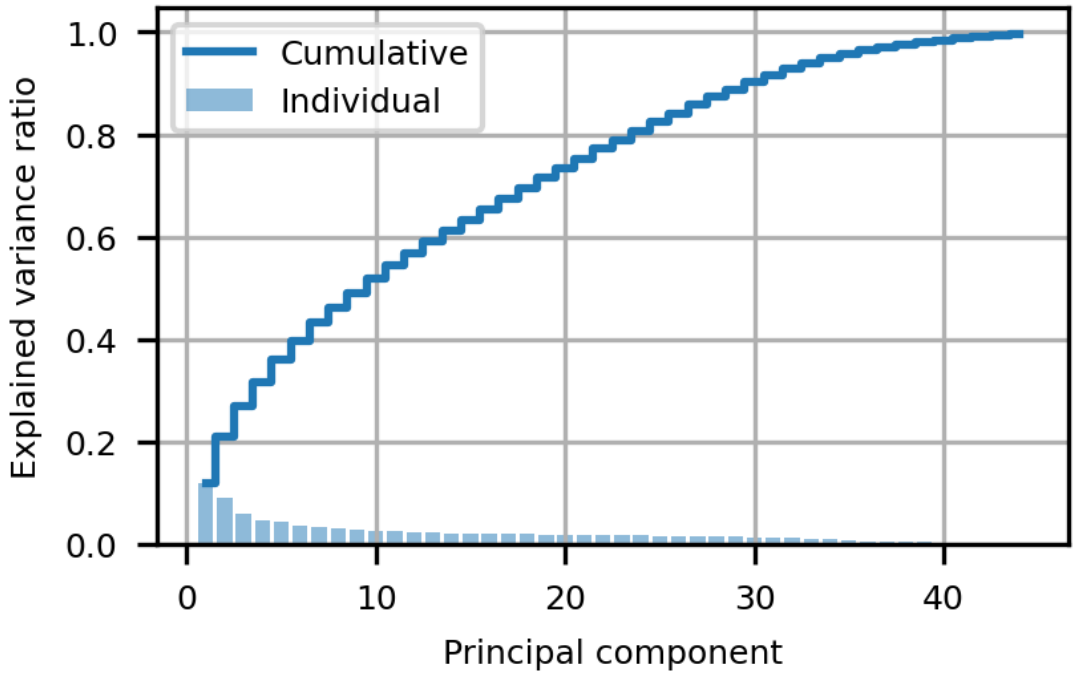


Figure 3.2: The explained variance ratio for the PCA. We see that all the variance is explained by 45 variables. Adding more variables will therefore not give any more information.

3.4.3 DNN

Different structures were tested but it was concluded after testing that 8 layers of 1000 neurons per layer could capture most of the complexity of the problem since it could overfit to around 95%, adding more neurons and layers did not increase the performance but after some point lowered it instead. To counter the overfit a dropout was added between each layer and was set to 0.01. The DNN was used on the PCA data.

3.4.4 1D convolutional

The 49 features were organized into five channels with 10 features in each. The channels were organized as shown in table 3.1. The order of each feature was tested to not have a significant effect as well as permuting elements from the different channels.

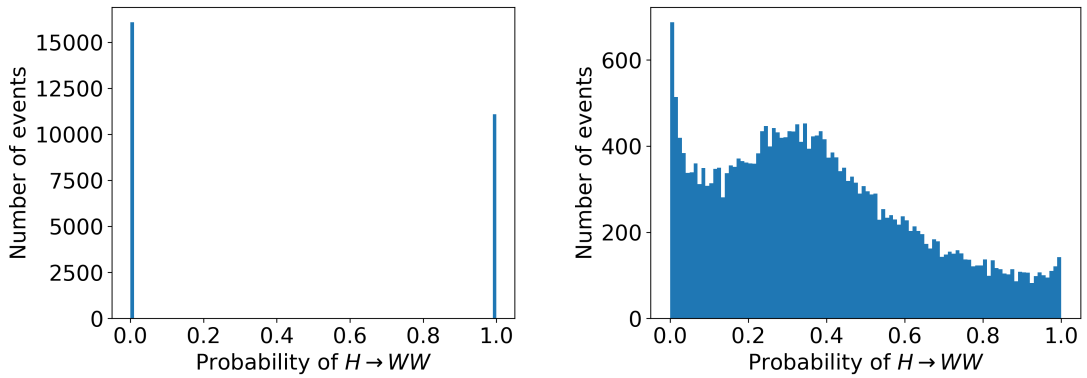
Table 3.1: The channels of the 1D convolutional model were made up of 49 features and a 0 element. They were organized so that information that might be related would be in the same channel.

1	2	3	4	5
Scalars and Zero element	Lepton 1	Lepton 2	Jets	

With two layers of convolution with 128 and 256 neurons with kernel size 3 and then one dense layer of 128 neurons it seems like the model could capture all of the complexity of the model since it could overfit the training data to 100%. A dropout of 0.2 was added to prevent this overfit. One layer of max pooling was added after the last convolutional layer to prevent overfit and increase the generalizability of the model.

3.5 Output

The output from the networks can be seen as a predicted probability of an event is either the labeled $H \rightarrow WW$ (1) or something else (0). The output of both MVA classifiers respectively when applied on the detector data can be seen in fig. 3.3, note that the 1D convolutional network gives more of a spectrum than the DNN. If applied on the training data a structure like fig. 3.3a would be expected while when using it on new data a structure more similar to fig. 3.3b is expected.



(a) DNN

(b) 1D convolutional network.

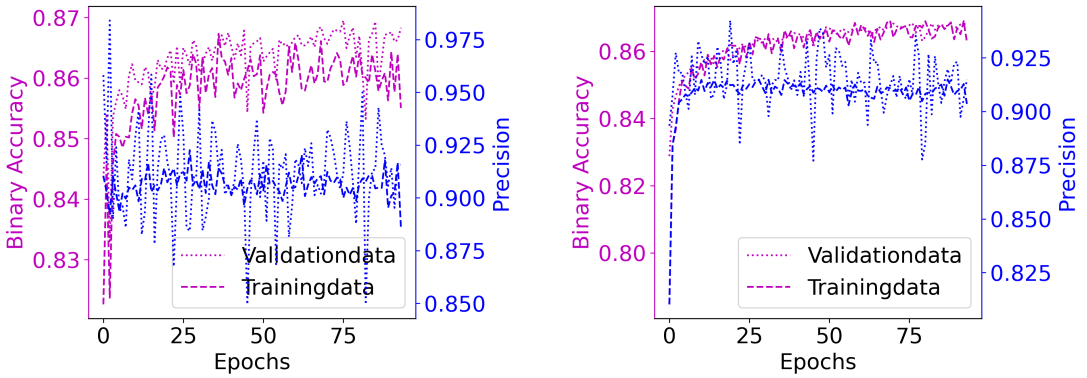
Figure 3.3: Output on the real data from the respective network. Note that the convolutional network gave more spectra that can be viewed as a probability of being either 0 och 1 while the DNN gave more of a labeling.

The metrics used to evaluate the models is the binary accuracy and precision on the training and validation sets which are both simulated data. They are defined as

$$BA = \frac{TP + TN}{FP + FN + TP + TN}, \qquad P = \frac{TP}{FP + TP} \tag{3.3}$$

where TP is the number of correctly predicted positives (1), TN is the number of correctly predicted negatives (0), FP is the number of incorrectly predicted positives and FN is the number of incorrectly predicted negatives.

The training and validation scores can be seen for each epoch in fig. 3.4. Worth noting is that the precision stays more or less constant while the accuracy is improving over the epochs. The model was saved for classification after the epoch with the lowest validation loss.



(a) DNN **(b)** 1D convolutional network

Figure 3.4: Training and validation plots. Both networks converge quickly and fluctuate a little, the fact that the validation score fluctuates significantly more is reasonable since the model is only trained on the training data and the validation data tests how general the model is. Note that the network was saved at the lowest validation loss for both.

4

Results

The results are divided into three parts. First the predicted results of the simulated data, then how well the neural networks could filter the data, and lastly the value of the CGLMP inequality.

4.1 Predicted results of simulated data

Based on the simulated data the inequality was calculated for only H→ WW and for random parts of the background with different proportions of Higgs and background. This shows that most of the background has a significantly lower value than the H → WW decay. The errorbars are the standard deviation scaled as it will be with 425 (the predicted number from section 3.1) events. The blue stripe is the predicted value by Barr and Fabbrichi et al. [8, 9].

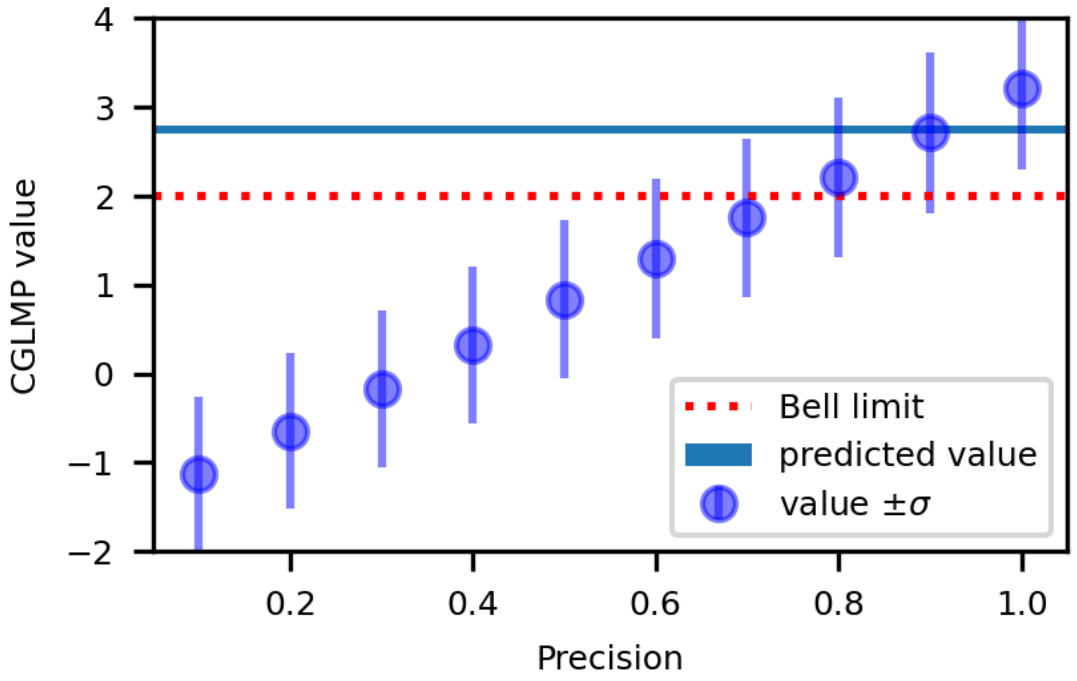
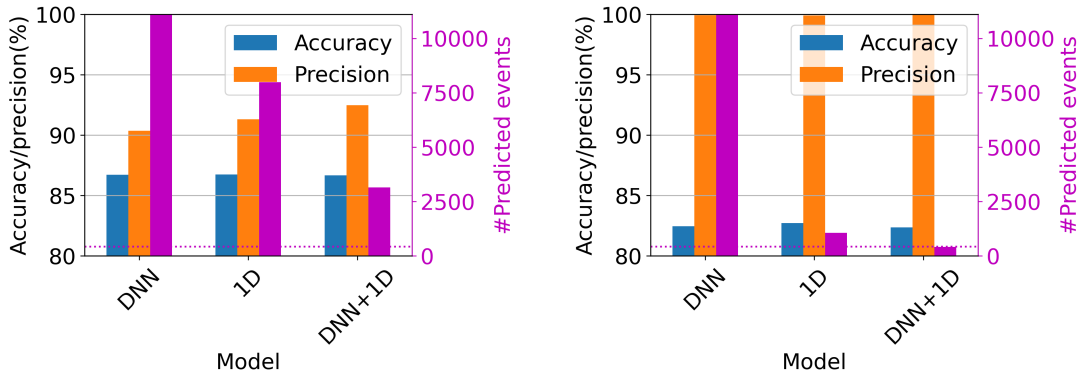


Figure 4.1: Bell value for different precisions. Events were picked randomly. The error bars are based on 425 events.

4.2 Performance of MVA

The DNN and 1D convolutional neural networks had a binary accuracy of 87% and precision of 90 and 91% each, by combining the two it rose to 87 and 92%. When the point where an event was classified as H→WW was changed to fit in with the number of predicted events the precision rose to 100% and the accuracy fell to 82%. The number of events passed were around 13000 for the DNN in both cases, for the 1D convolutional network decreased from 8000 to 800 and for the combination it decreased from 3500 to 425. A visualization of this can be seen in fig. 4.2.

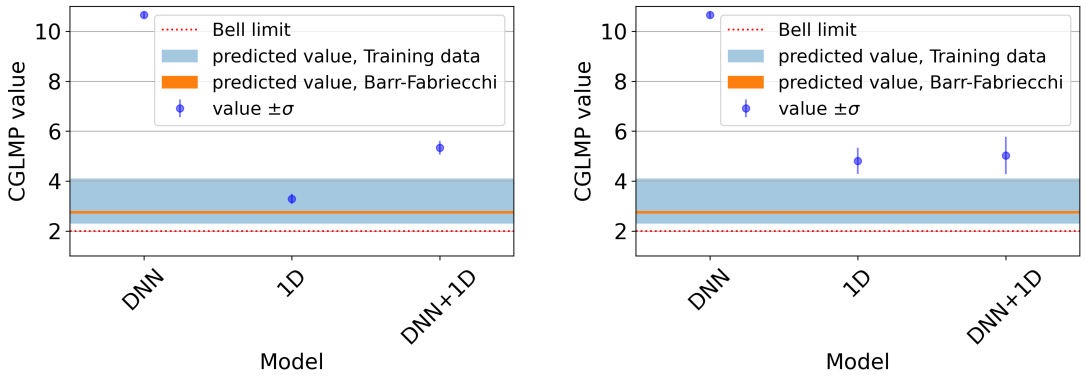


(a) With rounding at 0.5 **(b)** With rounding at 0.9

Figure 4.2: Accuracy, precision, and the predicted number of events for both models separately as well as combined for two different points of rounding. Note that the precision is close to 100% when the rounding is adjusted so that the models predict the correct number of events.

4.3 Bell value

Using the models described above to select the events predicted to be H → WW to calculate the CGLMP inequality we get the values 10.66 ± 0.12, 3.29 ± 0.2 and 5.34 ± 0.28 if we round at 0.5 and 10.66 ± 0.12, 4.81 ± 0.52 and 5.03 ± 0.75 if we round at 0.9 for the DNN, 1D convolutional and the both combined. As shown in figure fig. 2.3 can values between -20 and 30 be achieved and values over 2 break the classical limit where 2.9 means maximum entanglement for events from a three-level system as discussed in section 2.1.6.



(a) With rounding at 0.5 **(b)** With rounding at 0.9

Figure 4.3: The value for the CGLMP inequality calculated from eq. (2.7) with the error bar being calculated with eq. (2.8) and then scaled to be an error of the mean by eq. (2.9) for each model predicted events as well as the events predicted of both models combined.

5
Conclusion

Finally, I discuss the interpretation of these results and the probable causes of it. I also present some of the possible improvements and discuss how this could be done in future projects.

5.1 Discussion

By looking at the results in section 4.2 it seems like the 1D convolutional network alone can filter out the right amount of events that have almost only Higgs for the training data. The DNN seems to be less efficient with this since it does not give a spectrum of probabilities but labels to the data which means that for this type of network is the simulated and real detector data not any different. If they were just slightly different this should result in some uncertainty of the classifier where it gives a spectrum of probabilities of an event being either signal or background just as we see the 1D convolutional do. The combination of the two of them gives an even higher precision than the 1D convolutional alone even if the difference is slim and could be achieved by increasing the rounding point further.

When the CGLMP inequality is calculated in section 4.3 it seems like it is higher than the expected 2.8 which means maximum entanglement. It is 4.3σ over the classical limit 2 but also 3σ over the expected value. This might be because the neural networks do not perfectly select the Higgs events as it does on the training data and the background in all other variables looks similar but is not Higgs and just happens to give a high CGLMP value. Since many of the events do not come from the three-level system that characterizes the H→WW decay can these take values in a wide range between -20 and 30 which do not have to mean anything in terms of entanglement.

The assumption that the CM frame and the lab frame are the same or close to the same is also something that could be causing a distortion of the value. If there is a significant difference the angles would be more open and the CGLMP value would decrease as seen in fig. 2.3 and thereby come closer to the expected value. This might be the reason for the discrepancy between the simulation results I obtained and what has been previously seen.

5.2 Selection bias

Since the Higgs signal in the training data has angles that give the CGLMP inequality a value clearly over two, the MVA classifiers trained on this data will probably select data that has similar attributes such as angles. Therefore, this method will probably return values of the CGLMP inequality that has similar values to the training data. This means that if the simulated data is breaking the CGLMP inequality the MVA classifier will probably label the data that also breaks the CGLMP inequality since that is what it is trained on. It would therefore be good to exclude the angles from the data for the selection to leave the MVA more unbiased from that part of the data. It may also be reasonable to investigate how the results may change if the dimensions were reduced further. This will sacrifice some of the information but leave a simpler problem by using fewer dimensions.

5.3 Possible improvements

This thesis has used relatively simple methods as MVA classifiers. This seems to have worked to some extent but more advanced methods might be needed to reach the desired results. Some suggestions of these is to use t-Distributed Stochastic Neighbor Embedding (t-SNE) instead of PCA which is a more robust way of decreasing the dimension of the data [31]. One of the main differences between PCA and t-SNE is that t-SNE preserves small differences while PCA preserves large differences between pairs of data points.

It would also be possible to do feature testing by setting one feature to random values at a time to see how much the results changed and thereby determine the less important features to the classification task and then ignore them to decrease the dimensionality of the data [32].

A more advanced method for a MVA classifier that has worked well in Higgs classification earlier is boosted trees. This builds on decision trees and gradient boosting which has been proven successful for a few applications where Higgs classification was one [33].

5.4 Outlook

Already looking at the first part of the result we see that the data from the collider that is in the dataset just is not enough to get a result with much more than 1σ significance, to get to a 5σ level of significance around five times more data is needed. Some of this might already exist outside of the open dataset and more of it is currently produced at the LHC in run 3.

From this thesis report, it is apparent to me that the 49 variables that were used might not be enough to classify $H \longrightarrow WW^*$ which means that more information from the detector is needed to solve this task. This might be because of how the

open dataset is structured and the simplifications that have been made to make it easier to use. It might also be because the detector just is not good enough in run 2 where the data was taken.

Bibliography

[1] Wikipedia . [Online]. Available: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles_edit.svg

[2] A. N. Laboratory, "Simulation of large hadron collider events using leadership computing," May 2023. [Online]. Available: https://www.flickr.com/photos/argonne/22809522845

[3] G. Aad, B. Abbott, J. Abdallah, R. Aben, M. Abolins, O. AbouZeid, H. Abramowicz, H. Abreu, R. Abreu, Y. Abulaiti *et al.*, "Measurements of the higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 tev in the atlas experiment," *The European Physical Journal C*, vol. 76, no. 1, pp. 1–51, 2016.

[4] ATLAS, "Description of the 13 tev atlas open data branches and variables." [Online]. Available: http://opendata.atlas.cern/release/2020/documentation/datasets/dataset13.html

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

[6] 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, sep 2012. [Online]. Available: https://doi.org/10.1016%2Fj.physletb.2012.08.020

[7] J. S. Bell, "On the einstein podolsky rosen paradox," *Physics Physique Fizika*, vol. 1, no. 3, p. 195, 1964.

[8] A. J. Barr, "Testing bell inequalities in higgs boson decays," *Physics Letters B*, vol. 825, p. 136866, feb 2022. [Online]. Available: https://doi.org/10.1016%2Fj.physletb.2021.136866

[9] M. Fabbrichesi, R. Floreanini, E. Gabrielli, and L. Marzola, "Bell inequalities and quantum entanglement in weak gauge bosons production at the lhc and future colliders," 2023. [Online]. Available: https://arxiv.org/abs/2302.00683

[10] Wikipedia , "Particle physics," Mar 2023. [Online]. Available: https://en.wikipedia.org/wiki/Particle_physics

[11] D. Sternheimer, ""the important thing is not to stop questioning", including the symmetries on which is based the standard model," in *Geometric Methods in Physics: XXXII Workshop, Białowieża, Poland, June 30-July 6, 2013*. Springer, 2014, pp. 7–37.

[12] I. P. Ivanov, "Minkowski space structure of the higgs potential in the two-higgs-doublet model," *Phys. Rev. D*, vol. 75, p. 035001, Feb 2007. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.75.035001

[13] F. De Martini, "The higgs field and the resolution of the cosmological constant paradox in the weyl-geometrical universe," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 375, no. 2106, p. 20160388, 2017.

[14] N. A. Törnqvist, "Suggestion for Einstein-podolsky-rosen Experiments Using Reactions Like $e^+e^- \to \Lambda\bar{\Lambda} \to \pi^- p \pi^+ \bar{p}$," *Found. Phys.*, vol. 11, pp. 171–177, 1981.

[15] ATLAS Collaboration, "Introduction to jets." [Online]. Available: https://atlassoftwaredocs.web.cern.ch/AnalysisSWTutorial/obj_jet_intro/

[16] I. Burmeister, "Measuring the b-tagging efficiency in atlas data at $\sqrt{s}$= 8 tev and $\sqrt{s}$= 13 tev using the p rel t-method," 2017.

[17] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Physical review letters*, vol. 23, no. 15, p. 880, 1969.

[18] B. S. Cirel'son, "Quantum generalizations of bell's inequality," *Letters in Mathematical Physics*, vol. 4, pp. 93–100, 1980.

[19] J. J. C. Mckeown, "A Contribution to ATLAS Open Data Collaboration at CERN," 2019, presented 15 May 2019. [Online]. Available: http://cds.cern.ch/record/2704159

[20] C. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten, L. Lönnblad, S. Mrenna, S. Prestel, C. T. Preuss *et al.*, "A comprehensive guide to the physics and usage of pythia 8.3," *SciPost Physics Codebases*, p. 008, 2022.

[21] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual," *The European Physical Journal C*, vol. 72, no. 3, mar 2012. [Online]. Available: https://doi.org/10.1140%2Fepjc%2Fs10052-012-1896-2

[22] [Online]. Available: https://home.cern/news/news/accelerators/lhc-report-final-days-run-2

[23] Sartorius, "What is principal component analysis (pca) and how it is used?" Aug 2020. [Online]. Available: https://www.sartorius.com/en/knowledge/science-snippets/what-is-principal-component-analysis-pca-and-how-it-is-used-507186#:~:text=Principal%20component%20analysis%2C%20or%20PCA,more%20easily%20visualized%20and%20analyzed.

[24] A. R. Aouichaoui, R. Al, J. Abildskov, and G. Sin, "Comparison of group-contribution and machine learning-based property prediction models with uncertainty quantification," *31st European Symposium on Computer Aided Process Engineering*, p. 755–760, 2021.

[25] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *International Journal of Engineering Applied Sciences and Technology*, 2020.

[26] J. Brownlee, "1d convolutional neural network models for human activity recognition," Aug 2020. [Online]. Available: https://machinelearningmastery.com/cnn-models-for-human-activity-recognition-time-series-classification/

[27] D. Godoy, "Understanding binary cross-entropy / log loss: A visual explanation," Jul 2022. [Online]. Available: https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a

[28] [Online]. Available: https://optimization.cbe.cornell.edu/index.php?title=Adam#:~:text=Adam%20optimizer%20is%20the%20extended,processing%20in%20the%20future%20years.

[29] Open Data Atlas, "13 tev atlas open data." [Online]. Available: http://opendata.atlas.cern/samples-13tev/

[30] ——, "Atlas open data – development of a simplebut-real hep data analysis framework." [Online]. Available: https://cds.cern.ch/record/2709168/files/ATL-OREACH-PROC-2020-001.pdf

[31] A. GUPTA, "Pca vs t-sne," Jul 2020. [Online]. Available: https://www.kaggle.com/code/agsam23/pca-vs-t-sne

[32] A. Kumar, "Testing features of ml models - dzone," Aug 2018. [Online]. Available: https://dzone.com/articles/testing-features-of-ml-models

[33] T. Chen and T. He, "Higgs Boson Discovery with Boosted Trees," in *Proceedings of the NIPS 2014 Workshop on High-energy Physics and Machine Learning*, ser. Proceedings of Machine Learning Research, G. Cowan, C. Germain, I. Guyon, B. Kégl, and D. Rousseau, Eds., vol. 42. Montreal, Canada: PMLR, 13 Dec 2015, pp. 69–80. [Online]. Available: https://proceedings.mlr.press/v42/chen14.html

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY