Abstract: This report discusses the construction, performance, and use of a deep neural network to distinguish signal and background from a data set. The deep neural network is trained on the $VH \rightarrow q\bar{q}b\bar{b}$ resonance which is reconstructed from data collected by the ATLAS detector at the LHC running at 79.8 fb$^{-1}$ and $\sqrt{s} = 13$ TeV. The performance of the deep neural network is analyzed using the results from traditional cut based analysis.
Contents

1 Introduction 3
   1.1 A Crash Course on the Standard Model .......................... 3
   1.2 CERN, the LHC, and ATLAS ............................... 4

2 Reconstructing the $Y \rightarrow H + V$ Decay 6
   2.1 The $Y \rightarrow H + V$ Decay .................................. 6
   2.2 Using Jets to Compartmentalize Data .......................... 8
   2.3 Background Composition ...................................... 10

3 Identifying Signal and Background Data 11
   3.1 Preparing Signal and Background Data Sets ..................... 11
   3.2 Analysis of Signal Data from Cut Based Analysis ............... 13

4 Creating a Deep Neural Network 16
   4.1 A Brief Introduction to Machine Learning ...................... 16
   4.2 Creating the Deep Neural Network ............................. 16
   4.3 Standard Machine Learning Analysis of the Deep Neural Network .. 18

5 Using the Deep Neural Network with Data 20
   5.1 Signal Efficiency of the DNN .................................. 20
   5.2 Reconstructing $Y \rightarrow VH \rightarrow q\bar{q}bb$ Candidates ............ 20

6 Conclusions 23

7 Acknowledgements 24

References 25
1 Introduction

1.1 A Crash Course on the Standard Model

The atom is like a miniature Solar System. At the center, there is a massive nucleus full of positively charged protons and neutral neutrons. Orbiting the nucleus, like tiny subatomic asteroids, are negatively charged particles called electrons. Presently, it is believed that the electrons are elementary. However, protons and neutrons are not so simple. Like the Sun, they are much more complicated.

Protons and neutrons, referred to as hadrons, have been found to be composed of three even smaller particles called quarks. Six quarks are have been observed to date, creatively named by how strange they were when they were found or by the direction they spin in. Three quarks can be “glued” together by gluons, theorized particles that carry the Strong Force, to form hadrons like protons and neutrons. Particles that can combine to create hadrons are also referred to as partons. These particles can be seen in Figure 1. In this figure, in addition to the quarks, one can also see force particles, like the gluon, and leptons, like the electron and the muon. All six quarks and all six leptons also have an associated antiparticle, which are not shown in Figure 1 for the sake of compactibility.

Figure 1: Physicists’ present concept of the Standard Model and the graviton. [1].
Another important force involving quarks, leptons, and the Standard Model of particle physics is the Weak Force, which is carried by the $W^\pm$ and $Z^0$ bosons, referred to together as $V$ Bosons, and the Higgs Boson, referred to as an $H$ boson. As implied by the name, the Weak force is much weaker than the Strong force, as shown in Figure 2. The Weak force is also weaker than the Electromagnetic force, however it is still far stronger than the Gravitational force to the extent that the effect of the Gravitational force is negligible with regards to the interactions between sub-atomic particles. Both $V$ Bosons and Higgs Bosons have short lifetimes, but decay into particles with longer lifetimes that can be detected with modern day detectors. Using these decay products, the $V$ Bosons were discovered in the 1980’s but the Higgs Boson took until 2012 to discover in an experimental setup. This is because the Higgs Boson is so massive that it requires a particle accelerator strong enough to account for its mass. After 30 years of planning and building, the Large Hadron Collider was built.

### 1.2 CERN, the LHC, and ATLAS

CERN, standing for the European Organization for Nuclear Research\(^1\), is an incredible collaboration of scientists and engineers from institutions and laboratories all over the world who use their combined brain power to break through the boundaries of modern day science. Together, they built the largest particle accelerator on Earth. The Large Hadron Collider accelerates “bunches” of protons to near the speed of light and collides the bunches in various detectors on the ring of the Large Hadron Collider. Each detector has its own specialties and each associated collaboration has

\(^1\)The acronym was actually formulated from the French title Conseil Européen pour la Recherche Nucléaire. Overtime, the name was changed and translated into various languages but the acronym remains the same.
their own goals for what they’d like to discover with their detector. The data in this report comes from the ATLAS detector, shown in Figure 3.

ATLAS, A Toroidal LHC ApparatuS, is a general purpose composite detector situated in Meyrin, Switzerland designed to track and measure qualities of proton-proton collision products. ATLAS uses pixel detectors to track particles, calorimeters to measure the energy of particles, and muon chambers to detect and measure energy loss of muons [3]. Large toroidal and solenoidal magnets are used to bend charged particles from each bunch crossing so that their charge and transverse momenta may be measured. The information from all of the detectors can be used to calculate other qualities of reaction products such as mass and direction of motion. Presently, bunches of protons cross once every 25 nanoseconds. This translates to 40 million events every second, meaning the ATLAS detector has a lot of events and a lot of information to process.
# Reconstructing the $Y \to H + V$ Decay

## The $Y \to H + V$ Decay

The discovery of the Higgs Boson validates the Standard Model of physics in reference to particles that interact with energies up to a few hundred GeV. This validity expires when strong radiative corrections are applied to the Higgs boson’s mass, which implies a potential for the existence of new and undiscovered physics around the energy scale of the Higgs boson mass. Experts suspect that the new physics will involve substantial coupling to the Higgs boson, such as a particle decaying into a Higgs boson and another Standard Model particle, for example [4].

This idea is the basis for the resonance investigated in this report. As protons collide and quickly change into new particles within the ATLAS detector, the rapidly changing strong, electromagnetic, and weak forces combined with the rearrangement of quarks in the detector can manifest into fascinating particles. One such particle is the proposed $Y$ particle, which has been proposed to decay into a Higgs boson and a $V$ boson.

As mentioned in Section 1.1, the Higgs and $V$ bosons both have short lifetimes and must be measured by their decay products. Figure 4 shows the branching ratios (probability of each set of decay products) as a function of the mass of the Higgs boson. Considering the mass of the Higgs is 125 GeV, according to the figure, the most likely decay product is a bottom and an anti-bottom quark. From the branching ratios for $W$ boson decays shown in Table 1, it is clear that $W$ bosons most often decay into hadrons, meaning a quark and an antiquark. From Table 2 it is clear that $Z$ bosons most often decay into hadrons as well, implying another quark and antiquark pair.
Figure 4: Branching ratio for the Higgs boson decay [5].

<table>
<thead>
<tr>
<th>Mode</th>
<th>Branching Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e\nu$</td>
<td>$(10.71 \pm 0.16)%$</td>
</tr>
<tr>
<td>$\mu\nu$</td>
<td>$(10.63 \pm 0.15)%$</td>
</tr>
<tr>
<td>$\tau\nu$</td>
<td>$(11.38 \pm 0.21)%$</td>
</tr>
<tr>
<td>hadrons</td>
<td>$(67.41 \pm 0.27)%$</td>
</tr>
</tbody>
</table>

Table 1: Four important branching ratios for $W^\pm$ boson decays. [6]

<table>
<thead>
<tr>
<th>Mode</th>
<th>Branching Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ee$</td>
<td>$(3.3632 \pm 0.0042)%$</td>
</tr>
<tr>
<td>$\mu\mu$</td>
<td>$(3.3662 \pm 0.0066)%$</td>
</tr>
<tr>
<td>$\tau\tau$</td>
<td>$(3.3696 \pm 0.0083)%$</td>
</tr>
<tr>
<td>Invisible</td>
<td>$(20.000 \pm 0.055)%$</td>
</tr>
<tr>
<td>Hadrons</td>
<td>$(69.911 \pm 0.056)%$</td>
</tr>
</tbody>
</table>

Table 2: Four important published branching ratios for $Z^0$ boson decays [6].
Therefore, the entire assumed reaction is that two protons collide to create the $Y$ particle. This $Y$ particle then decays into one Higgs boson, which then decays into a bottom quark and an antibottom quark, and into one $V$ Boson, which then decays into a quark and an antiquark. This entire proposed reaction is drawn in Figure 5.

![Feynman Diagram](image)

Figure 5: Feynman Diagram of the proton proton collision resulting in a $Y$ particle that decays into a Higgs and a $V$ Boson which then (most often) decay into bottom quarks and quarks respectively.

### 2.2 Using Jets to Compartmentalize Data

In any given event, there is noise from previous events, called pileup, noise from interactions that are not being investigated, and signal from the investigated interaction of the crossing proton bunches. The anti-$k_t$ algorithm is used to classify important data as jets [4]. A jet is a cone shaped region in the detector, originating at the interaction point, which encompasses the detectors that received important signals from an event. A large-radius (large-R) jet encapsulates a large region of the detector and can be broken down into subjets via the $k_t$ algorithm and trimmed to reduce extraneous noise [4]. The subjets are smaller cones and can therefore represent individual particles more accurately.
All jets are characterized by the mass that they contain \((m)\), the transverse momentum with which they travel away from the beamline \((p_T)\), and the pseudorapidity \((\eta)\) and azimuthal angle \((\phi)\) detailing their direction. The coordinate system involving pseudorapidity and azimuthal angle is shown in Figure 6. The size of large-R jets is defined in \(\phi - \eta\) space as \(R = 1.0\), where \(R\) is defined in Equation 1. Large-R jet structure can be described by the D2 substructure variable which is used at the 50\% efficiency working point [8]. The subjets are of size \(R = 0.2\) and are associated with a multivariate discriminant called the MV2c10. This variable helps to identify quarks that are the result of a particle decaying into bottom and/or antibottom quarks (b-jets) and is especially important when determining if a large-R jet represents a Higgs boson [9]. Figure 7 shows the MV2c10 output for various types of jets and is used to consider how the MV2c10 cut should be defined when tagging b-jets.

\[
R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}
\]  

(1)

The aforementioned variables are considered to be low level variables in the context of this report. If one were to use the low level variables to create new variables, such as the mass of a reconstructed \(Y\) particle jet, those would be considered high level variables. The very low level variables, such as the raw timing and signal data from the ATLAS detector, will not be considered.
Figure 7: Output of the MV2c10 boosted decision tree (BDT) for three types of jets, including b-jets [9].

### 2.3 Background Composition

The background involved in this study includes pileup and interactions that are not the focus of this study. Pileup is avoided by only considering large-R jets with a transverse momentum greater than 200GeV and by using only the leading and sub-leading large-R jets from each event. Multijets attribute to 90% of the background from significant interactions that this study is not interested in, followed by $t\bar{t}$ events which represent less than 10% and $V +$ jets events which dominate the small remainder. Multijets prove to be difficult as they tend to resemble the $Y$ particle in regards to transverse momentum and pseudorapidity. To prevent against observing signal that does not correspond to the $Y$ particle, such as other $VH$ decay forms and multijets, various cuts can be performed on the data and a lepton veto can be established [4].
3 Identifying Signal and Background Data

3.1 Preparing Signal and Background Data Sets

Each large-R jet in the ATLAS detector usually corresponds to one major particle, like the Higgs Boson or a $V$ Boson. Some of the low level data from the ATLAS detector can be used to identify the type of particle a large-R jet represents. Considering this study is interested in the nature of a decay into a Higgs and $V$ Boson, the events that correspond to this possible two body decay are isolated from the background events that do not represent the decay. Every event in a data set is cut based on the information in Table 3. The terms “leading” and “subleading” refer to the large-R jets with the highest and second highest transverse momenta in a single event, respectively. The collection of events that pass these cuts are referred to as the preselection events. The lepton veto mentioned in the previous section and the leptons mentioned in Table 3 are determined based on the criteria in Table 4. The $HV \rightarrow q\bar{q}b\bar{b}$ final state that is investigated does not include leptons, this is referred to as an all-hadronic final state, so if any jets pass the criteria in Table 4 then the event is rejected and regarded as background.

<table>
<thead>
<tr>
<th>Quality</th>
<th>Cut</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two large-R Jets, each with Mass $</td>
<td>\eta</td>
</tr>
<tr>
<td>$p_T$ of Leading Jet $\geq 450$ GeV</td>
<td></td>
</tr>
<tr>
<td>$p_T$ of Subleading jet $\geq 250$ GeV</td>
<td></td>
</tr>
<tr>
<td>Number of Leptons $0$</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Preselection criteria for events that may be the result of the $Y$ particle decay [4].

In addition to the criteria in Table 3, Table 5 describes the criteria with which a code can determine if a large-R jet is considered to be a Higgs Boson or a $V$ Boson. If the jet contains bottom or anti-bottom quark subjets, the large-R jet is referred to as being $b$-tagged once per subjet that passes the MV2c10 cut. The MV2c10 cut was defined with the MV2c10 BDT outputs from Figure 7 in mind, and it was determined that subjets with an MV2c10 value of 0.3706 or higher would be regarded as $b$-tagged given the high ratio of $b$-jets with MV2c10 values at and above this value. It should be noted that the error associated with mis-tagging a subjet is much smaller relative to the error associated with incorrectly classifying a multijet event as a preselection event. If a large-R jet’s D2 value indicates a 2-prong substructure, then this large-R
Table 4: The criteria for classifying leptons in an event, which are used to reject events that are not solely hadronic. $d_0$ is an impact parameter perpendicular to the beam line and $\sigma_{d_0}$ is its error. $z_0$ is the distance along the beamline from the interaction point to the point where $d_0$ is measured [4].

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Cut</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transverse momentum</td>
<td>$&gt; 7$ GeV</td>
</tr>
<tr>
<td>$</td>
<td>\text{Pseudorapidity}</td>
</tr>
<tr>
<td></td>
<td>Muons: &lt; 2.5</td>
</tr>
<tr>
<td>$</td>
<td>d_0/\sigma_{d_0}</td>
</tr>
<tr>
<td></td>
<td>Muons: &lt; 3</td>
</tr>
<tr>
<td>$z_0 \sin (\theta)$</td>
<td>&lt; 0.5 mm</td>
</tr>
<tr>
<td>Isolated from other leptons</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 5: Criteria used to classify a jet as a Higgs Boson or a $V$ Boson.

<table>
<thead>
<tr>
<th>Boson</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higgs Boson</td>
<td>$95$ GeV $\leq$ Mass $\leq$ $145$ GeV At least one subjet with MV2c10 $\geq 0.3706$</td>
</tr>
<tr>
<td>V Bosons</td>
<td>$65$ GeV $\leq$ Mass $\leq$ $105$ GeV D2 at $50%$ Efficiency [8]</td>
</tr>
</tbody>
</table>

jet has the appropriate subjet structure to indicate that it could correspond to a $V$ Boson.

If a preselection event has both a Higgs and a $V$ Boson, the event is referred to as a signal event. Figure 8 provides a visual for which events are considered to be signal events based on the qualities of a large-R jet if a $V$ boson is present and no leptons are present. For these events, the 4-vectors of the leading Higgs and $V$ Boson jets can be added to investigate what type of particle they both decayed from, effectively reconstructing the $Y \rightarrow VH \rightarrow q\bar{q}b\bar{b}$ decay. The four vectors are composed of the pseudorapidity, mass, azimuthal angle, and transverse momentum of each large-R jet, and can be added to reconstruct the two body decay, as per conservation of 4-momentum. The result will reveal the mass, transverse momentum, and direction of motion of the unknown $Y$ particle that originally decayed.
Figure 8: Classification of which events are considered to be signal or background based on whether or not at least one of the large-R jets in the event has a particular mass or composition of $b$-tagged subjets, a $V$ boson present in the event, and no leptons present in the event. Jets that are classified in regions not shown in this graph are all considered background unless they have more than 2 $b$-tagged jets and fall into the 95 GeV to 145 GeV mass range.

3.2 Analysis of Signal Data from Cut Based Analysis

The methods in the previous section were implemented on a 44,800 event data set created using Madgraph5 [10] and Pythia [11] and on a 24,800,000 event background set compiled from 2015, 2016, and 2017 ATLAS data collected at $79.8\ fb^{-1}$ and $\sqrt{s} = 13\ TeV$. Every event in the simulated 44,800 event data set passes the criteria in Table 3 and is referred to as the preselection set. Using the criteria in Table 5 it was determined that of the 92,000 large-R jets present, 43,000 were $V$ bosons and 33,000 were Higgs Bosons. Figure 9 shows the mass, pseudorapidity, transverse momentum, and azimuthal angle distributions for all large-R jets in the preselection set and for the Higgs and $V$ boson jets, specifically. Of those 48,000 preselection events, 24,800 events had both a $V$ Boson and a Higgs boson, forming the signal set.

For each event in the signal set, 4-vectors were created from the leading Higgs and $V$ boson large-R jets, then added to calculate the initial variables characterizing the $Y$ particle. The plots in Figure 10 show the results of this cut based analysis. It is clear from these plots that a particle traveling along the beamline with a mass of around 2.8 TeV will be expected when the deep neural network is used to make predictions on data later in this report. Traditional cut based analysis has been shown to properly identify 22% of these signal events [4].
Figure 9: The distribution of variables characterizing all large-R jets (in black) in the preselection data set and those which passed the cut for V Bosons (in blue) and Higgs Bosons (in red) as a result of the selection criteria in Table 5.
Figure 10: The distribution of variables describing the $Y$ particle as a result of the reconstruction of the $Y \rightarrow VH \rightarrow qq\bar{b}\bar{b}$ decay process using cut based analysis.
4 Creating a Deep Neural Network

4.1 A Brief Introduction to Machine Learning

Artificial neural networks (NNs) begin as layers of neurons, called nodes, that have uniform weights of importance. Similarly to a student using flashcards to study for an exam, the NN is given data that has been identified as signal or background and uses this to train itself. The provided data set is split into a training set and a validation set where the former is used for large adjustments to the NN and the latter is for fine tuning [12]. Over the span of one epoch, the training set is passed through the NN so that it can adjust its weights via an activation function, then the validation set is passed to fine tune the changes. The training continues until the neural network does not improve any further over the span of multiple epochs, at which point the code declares that the NN is trained and ready to use. Leftover, never-before-seen data is then brought forward and passed through the NN so that its performance can be analyzed.

4.2 Creating the Deep Neural Network

The NN created for this study is a Deep Neural Network trained on the low level variables discussed in Section 2.2. This was chosen because Deep Neural Networks (DNNs) trained on low level variables have been proven to perform better than Shallow Neural Networks trained on both high and low level variables [13]. DNNs have even been proven to learn the high level variables and then some [13]. In addition, the DNN was trained from data scaled to fit a gaussian of mean 0 and a width of approximately 1, because scaled data sets usually train NNs more accurately and more efficiently than unscaled data sets [12]. The scaling was performed using scikit-learn's Standard Scaler feature [14].

The process begins with the 44,800 event preselection data set, created according to the cuts in Table 3, and the left-over 24.8 million event background data set. Using ATLAS software written in C++ and ROOT [15], both data sets were converted into a ROOT tree containing, for each event, the transverse momentum, mass, azimuthal angle, pseudorapidity, and D2 substructure variable of the two large-R jets with the largest transverse momentum, if the momentum is greater than 200 GeV. The tree also contains the MV2c10 values of the two subjets with the highest transverse momentum associated with the two large-R jets. The ROOT tree is then loaded into an external python script and can be used to create the DNN or to predict what events are signal if the DNN has already been created.
The DNN generated in this investigation has 16 input nodes, 3 hidden layers with 32 nodes each, and 1 output node. This DNN is shown in Figure 11. Using CERN’s Service for Web based ANalysis (SWAN), the DNN was created and trained with the Keras Python package [17] using the TensorFlow [18] backend and the RELU activation function\(^2\) on a computer using 4 cores and 10 GB of memory. The code loaded all necessary files and trained the DNN over 77 epochs in three minutes, characterizing a data set with almost 90,000 events that was composed of half preselection events and half background events. In ten minutes, on the same SWAN server, a separate script loaded the new NN and made predictions on almost 25 million events. Six minutes later, the predictions had been used to reconstruct the two body decay from the event and create graphs of transverse momentum, mass, azimuthal angle, and pseudorapidity of the reconstructed jet.

\(^2\)The REctified Linear Unit (RELU) activation function is defined as \(f(x) = \max(0, x)\).
Figure 12: The test accuracy and loss for the DNN created for the identification of signal and background for this two body decay which occurred within ATLAS.

4.3 Standard Machine Learning Analysis of the Deep Neural Network

The DNN tested very well after it was trained, implying that it should be a good resource to differentiate data between preselection and background events. Figure 12 shows the test loss and test accuracy of the neural network as it was being trained. The test accuracy smoothly improved as the network was trained, ending at 0.986 and the test loss behaved similarly, ending with a value of 0.041. This is very good and bodes well for the network. The Receiver Operating Characteristic (ROC) curve, Figure 13, shows a nice and steep rise followed by a sharp flattening of the curve, with an area underneath the curve of 0.999. This figure shows that when the testing data set was used, the DNN predicted the output very accurately.

To make a prediction, the neural network reads in the information from an event and outputs a number on a scale of 0.0 to 1.0, where 0.0 means the event is background and 1.0 means the event is a preselection event. Because the DNN is not exactly perfect, the predicted values are not just 0.0 and 1.0, but a range from 0.0 to 1.0. Figure 14 shows the distribution of prediction values for the testing data set. This histogram looks good because it is steep at the boundaries and flat in the middle, implying that most of the data was binary and only a small fraction of events were in between 0.0 and 1.0. Table 6 shows how the percentage of the data set which is predicted as a preselection event changes as the lower limit for the predicted value is changed. From this table it is satisfying to see that for a testing set that was half composed of preselection events, 50.3% of events were predicted to be preselection events when a preselection event was considered to be 0.5 to 1.0.
Figure 13: The ROC curve for the DNN created to differentiate two body decay signal and background from ATLAS data.

![ROC Curve for DNN](image)

ATLAS Work In Progress
\[ \sqrt{s} = 13 \text{ TeV} \]
\[ \int L = 79.8 \text{ fb}^{-1} \]

Full Curve (Area = 0.999)

Figure 14: Distribution of prediction values when the DNN was tested on the testing set.

![Distribution of Prediction Values](image)

<table>
<thead>
<tr>
<th>Lower Limit</th>
<th>Preselection Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>0.01</td>
<td>0.542</td>
</tr>
<tr>
<td>0.05</td>
<td>0.524</td>
</tr>
<tr>
<td>0.25</td>
<td>0.510</td>
</tr>
<tr>
<td>0.50</td>
<td>0.503</td>
</tr>
<tr>
<td>0.75</td>
<td>0.495</td>
</tr>
<tr>
<td>0.95</td>
<td>0.468</td>
</tr>
<tr>
<td>0.99</td>
<td>0.428</td>
</tr>
<tr>
<td>1.0</td>
<td>0.017</td>
</tr>
</tbody>
</table>

Table 6: Fraction of data predicted as preselection events as a function of the lower limit for preselection data prediction values.
5 Using the Deep Neural Network with Data

5.1 Signal Efficiency of the DNN

The signal efficiency is defined as the fraction of events which are predicted correctly. This is defined in Equation 2 and is equal to 1.0 if the DNN performs perfectly. In Section 3.2 it was noted that traditional cut based analysis had a signal efficiency of 22% [4].

\[
\text{Signal Efficiency} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Number of Events}} \quad (2)
\]

Using a set of pre-identified data, the DNN’s signal efficiency was calculated as a function of the lower prediction value limit and is shown in Figure 15 and Table 7. This is calculated only using signal data, so True Negative is always 0. From the table it is told that when the lower limit is set to 0.5, the signal efficiency is 99.1%. It should also be noted that even as the lower limit approaches 1.0, the signal efficiency is still higher than in cut based analysis by greater than 60%.

5.2 Reconstructing \( Y \rightarrow VH \rightarrow qqbb \) Candidates

For each event with a prediction value greater than 0.5, including background and preselection events, a reconstruction was performed using 4-vectors formed from the two stored large-R jets in the event, exactly like the reconstruction described in Section 3.2. Figure 16 shows the pseudorapidity, transverse momentum, and...
azimuthal angle distributions resulting from this reconstruction. By comparing these figures back to Figure 10, it is clear that these three distributions do not change very much from the cut based analysis.

When choosing 0.5 as the lower limit for what events are considered to be preselection events by the DNN, 1.7% of the background events are considered to be signal. This is low, which is good. However, there are 24,800,000 events in the background which corresponds to over 400,000 events incorrectly flagged as a preselection event, almost 10 times the amount of actual preselection events. The black curve in Figure 17 shows the mass distribution of the reconstructed jets from every single event tagged as a preselection event by the DNN. The blue curve in Figure 17 shows the distribution of background events that were predicted to be preselection events and the red curve shows the preselection events that were predicted to be preselection events. The former distribution peaks around 2200 GeV and the latter peaks around 2800 GeV, just as in Figure 10a. The peak from the preselection events is prominent in the overall black curve, however if the preselection and background peaks were closer together, this may not be so easy to detect.

Figure 18a shows the mass distribution of the resulting jet from the $Y \rightarrow H + V$ reconstruction performed on every event in the entire data set, before being passed through the DNN. This graph is overlaid with the mass distribution for reconstructed jets from only preselection events, in red, and from only background events, in blue. The blue, background curve lays almost precisely on top of the total event curve, and the red, signal curve is almost completely flat. For this reason, Figure 18b shows the reconstructed mass distribution for just the signal events. Clearly, the amount of background events in the sample has a large effect on the use of the DNN outputs.

Figure 16: Pseudorapidity, transverse momenta, and azimuthal angle distributions of the reconstructed Y particle for background events that were tagged as signal by the DNN. Notice that they are very similar to those in Figure 10.
Figure 17: Mass distribution of reconstructed jets from signal data tagged as signal, background data tagged as signal, and the sum of those two histograms to simulate real results.

and may pose issues in future analyses. However, the DNN created can effectively identify data that could be from the $Y$ particle reconstruction thus taking all the reconstructions shown in Figure 18a and creating a more exclusive distribution like in Figure 17.
(a) Mass distribution for reconstructed jets from every event in the preselection data set, background data set, and combined.

Figure 18: Mass distribution of the reconstructed jet when every event in the background and preselection data set have a reconstruction performed on the two leading jets.

(b) Mass distribution of reconstructed jets from all preselection events.

6 Conclusions

This purpose of this project was to construct and train a deep neural network to identify signal and background in a data set. The signal used for training was the $VH \rightarrow q\bar{q}b\bar{b}$ resonance which was simulated based on data collected by the ATLAS detector.

The created deep neural network tested very efficiently in development, achieving an ROC with an area under the curve value of 0.999. It was built within a few minutes and can analyze a data set with millions of events in a matter of minutes on a common computer. In performance, the DNN achieved a signal efficiency of 99.1%, correctly classifying almost all of a set of preselection data. Compared to the 22% signal efficiency of cut based analyses, the DNN proves to be a sound option for future analyses.

However, considering the $Y \rightarrow VH \rightarrow q\bar{q}b\bar{b}$ decay is uncommon, the effect of background events which were incorrectly predicted to be a result of this decay dominated over the effect of the true signal events. The background event reconstruction of an imposter $Y$ particle mass was around 2.2 TeV and the preselection data true $Y$ particle reconstruction peaked around 2.8 TeV, just like in the cut based analysis. The cut based analysis, while having a low signal efficiency, revealed a clear and isolated peak in the reconstructed $Y$ particle mass distribution. Considering the mass distribution created from the deep neural network results had two peaks, this method
may not always accurately reveal the existence of an unknown particle and should be considered carefully and potentially in conjunction with the cut based analysis to ensure a well rounded analysis.

7 Acknowledgements

I would like to thank Dr. Kalliopi (Poppy) Iordanidou for all of her guidance, mentorship, and assistance in every aspect leading up to this report. I would also like to thank Professor John Parsons, the administrators, and the organizers of the Nevis Laboratories REU and the National Science Foundation for making this opportunity possible. Lastly, I would like to thank CERN for successfully building, running, and maintaining the LHC and the ATLAS collaboration who have made this physics possible.
References


