The search for dark matter in final state events with one lepton and two b-jets at the ATLAS detector

James Tipping
201162824

UNIVERSITY OF LIVERPOOL

A thesis in partial fulfilment of the requirements for the degree of
Bachelor of Science (Hons)
Under the supervision of Monica D'Onofrio
At the Department of Physics
May 2019

Table of Contents

Declaration

I hereby declare that all work submitted is my own, having not been submitted previously for any award. If information has originated from a source other than me, their work has been acknowledged.

Acknowledgements

I would like to thank my parents for supporting me throughout my time at university, both practically and mentally. Without their love and encouragement, I would not have been able to do this. Also, I am very thankful to Monica for her help and guidance throughout this study. Her passion for what she does and devotion towards particle physics is something to be admired.

Abstract

The aim of this study is to search for evidence of dark matter candidates, in the context of supersymmetric models. The dark matter candidates are produced as the decay products of heavier supersymmetric particles. The events investigated are characterised by the presence of one lepton and a neutrino, two bottom quarks and the two dark matter candidates. The latter, as well as the neutrino, are identified as missing transverse momentum in the ATLAS detector. The background mimics the signal in data, producing two bottom quarks and a range of leptons, quarks and neutrinos. Measurable variables from these collisions are analysed for differences between signal and background using traditional cut-and-count techniques, and with machine learning (ML) to investigate its effectiveness with regards to multivariate analysis.

Introduction

The existence of dark matter is proven by several cosmological measurements, such as the way galaxies spin. It was found that the rotation speed of stars on the outside of galaxies was not slower, but the same as or faster than stars at the centre of the galaxy [12], defying Newton's laws of motion. This implies the mass of a galaxy is not centred in the middle, as previously thought, but is spread more evenly across the galaxy. However, since the majority of the matter we can observe is at the centre of the galaxy, another type of matter, one that is not observable by us, must be present in the outer regions of the galaxy, which causes the stars in the outer regions to spin more quickly [2] . It is estimated that there is 6 times more dark matter than visible matter in our universe.

The identity of this dark matter is highly sought after. One such candidate was the neutrino, found in the standard model of particles, which before the 1980's was a very viable candidate [8]. It is a very weakly interacting particle with mass, found in abundance such that they could explain all of the unobserved matter in the universe. Two types of dark matter were considered - hot (relativistic) and cold (non-relativistic). Since standard model neutrinos are very light particles, they are expected to begin in the early universe with relativistic speeds, thus being hot dark matter. However, therein lies a problem. Hot dark matter has been shown by numerical simulations to produce super high mass clusters (much higher mass than a galaxy), which then break up over time to form smaller structures such as galaxies. Our knowledge of the universe shows that in fact, smaller mass structures formed first, which then accumulated mass to produce larger structures. Formation in such a way indicates that the majority of dark matter must be cold, which rules out the standard model neutrino as a viable candidate [13]. Once neutrinos had been discarded as possible dark matter, it became apparent the Standard Model of particle physics, which describes matter and forces, did not include a suitable candidate. Weakly interacting massive particles (WIMPs) are one of the proposed candidates for dark matter.

Supersymmetry (SUSY) proposes that each boson has a supersymmetric fermion partner (with spin differing by ½) and vice versa. The lightest supersymmetric particle (LSP) is a good candidate for dark matter, as it is both stable and neutral (so doesn't decay further), meaning it is the best candidate for dark matter. As well as this, SUSY unites the three subatomic forces (strong, weak and electrostatic), and helps to explain why we observe the Higgs boson as being so light [7]. None of these supersymmetric particles have been discovered as yet, but

experiments are ongoing at the Large Hadron Collider (LHC), and efforts are continuing to search for these particles.

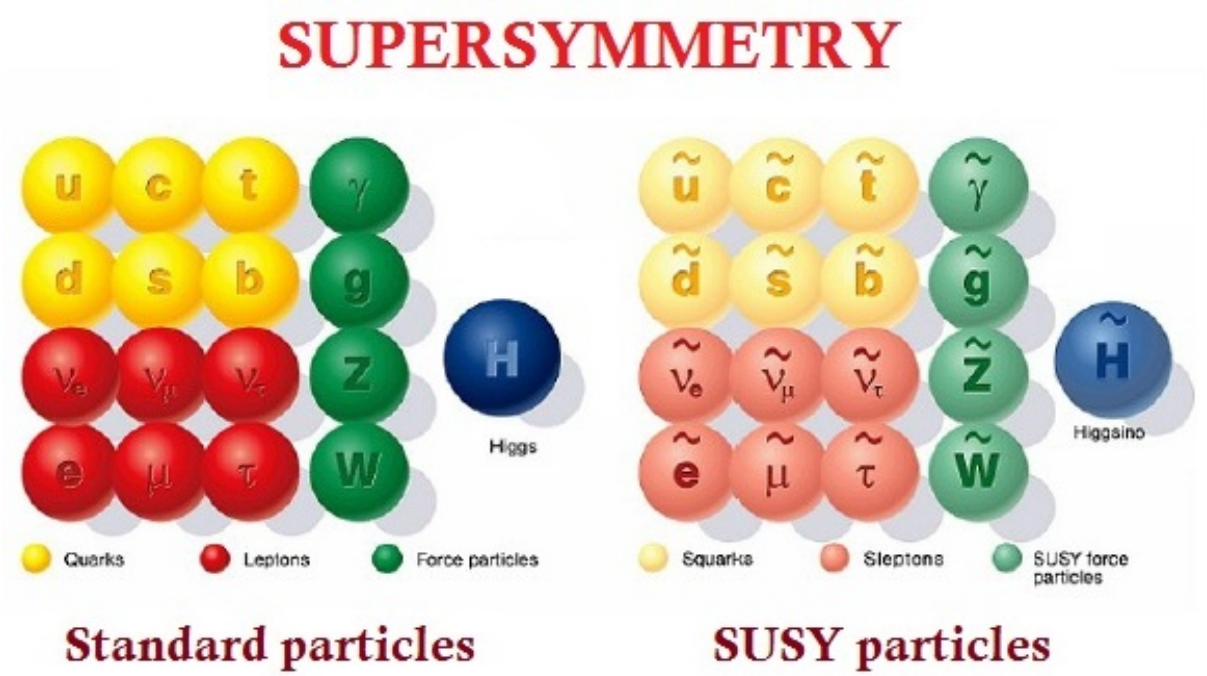


Figure 1 – The proposed supersymmetry model, which goes beyond the Standard Model of particles [11]

Amongst the various SUSY processes that could be searched for at the LHC, there is the production of charginos ($\tilde{\chi}^{\pm}$) and neutralinos ($\tilde{\chi}^{0}$), which are collectively referred to as electroweakinos. This is shown in figure 1. The neutralino with a multiplicity of 1 ($\tilde{\chi}_1^{0}$) is the LSP which was the focus of this study. Since the LSP is very weakly interacting, its presence cannot be observed by any instruments at the LHC. However, since under this hypothesis a Higgs and W boson are produced in conjunction with the LSP, and further decay into fermions, the momentum of the fermions can provide valuable information. Since the total transverse momentum should be zero, the 'missing' momentum in the detector arises from otherwise invisible particles – which could be the neutralinos – so their existence might be inferred by inspecting the momentum of the visible particles. This is called the missing transverse momentum (MET), which is an important quantity when trying to differentiate between events potentially producing supersymmetric particles and events which do not. [1]

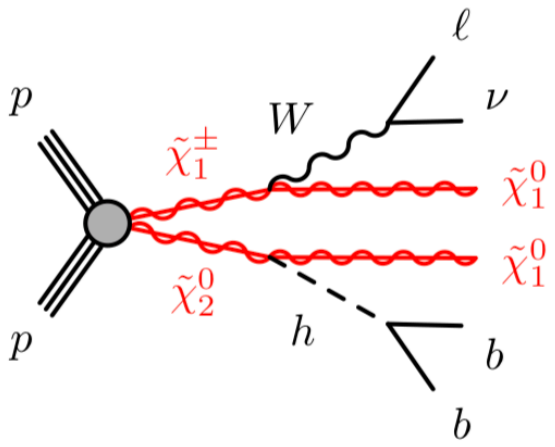


Figure 2 – The supersymmetric particle production method focussed on in this study, producing one lepton and two bottom quarks [1]

Motivation for this study

A previous study was performed by the ATLAS collaboration, considering various hypotheses for the masses of the particles. A range of possible masses for the chargino and neutralino were ruled out. Supersymmetric particles can be produced by a range of different decay processes, and these are covered in the previous study. The production method this study focused on is the one shown in figure 2, where two b-jets, a neutrino, lepton and two LSPs are produced. It is interesting to note that the next-to-lightest neutralino decays into a Higgs boson and the LSP. The Higgs boson decays into 2 b-jets around 60% of the time making this a good decay process to analyse. [1]

As mentioned, a range of possible masses for the neutralino and chargino have already been excluded. See below the graph showing the results of the previous study. Any combination of masses under the red line have already been excluded, and thus are not viable. Two mass combinations were tested in this study. A low mass sample consisted of a chargino and neutralino 2 with mass 300GeV, and a neutralino 1 mass (the LSP) of 150 GeV. A high mass sample consisted of a chargino and neutralino 2 mass of 800 GeV, with the neutralino 1 having a mass of just 1 GeV. Since the combination of masses of the particles cannot lie to the left of the grey dotted line, these two models were the most ‘extreme’ examples, to test if these analysis techniques worked in different signal regions. [1]

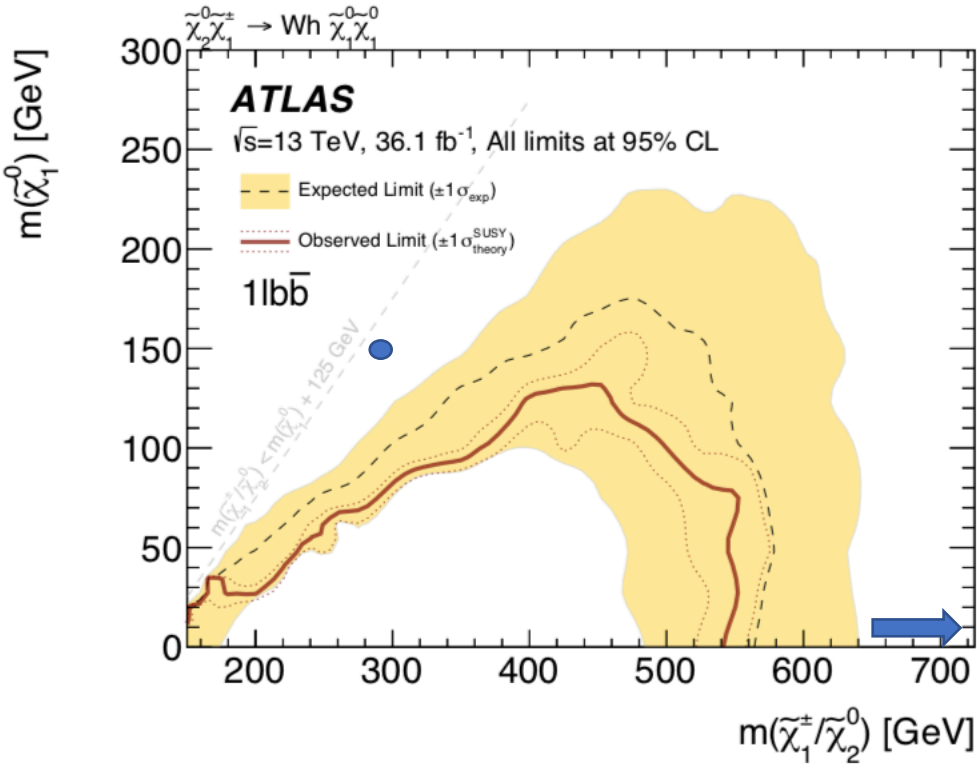


Figure 3 – The combination of masses for the chargino and neutralino carried out in the previous study. Masses below the red line have been ruled out as viable. The blue point shows the combination of masses for the low mass sample, as does the blue arrow for the high mass sample, which lies off the scale of the graph. [1]

The ATLAS detector

The ATLAS detector is a multi-purpose detector at the LHC. Particles collide at the centre of the detector, creating new particles, which then escape in all directions. There are 6 different detections systems present, which measure and record the particles direction, energy and momentum [4]. This is especially useful when trying to infer the presence of particles that we cannot detect. Inspecting the diagram below, it can be seen how exactly the presence of a particle can be inferred. Particle collisions in the ATLAS detector obey the principle of conservation of momentum, and thus the total transverse momentum before the collision must equal the total transverse momentum after – which is equal to zero, since the particles collide head on in the same plane. Therefore, if the total transverse momentum after the collision is not equal to zero, the momentum of the 'missing' particle can be calculated.

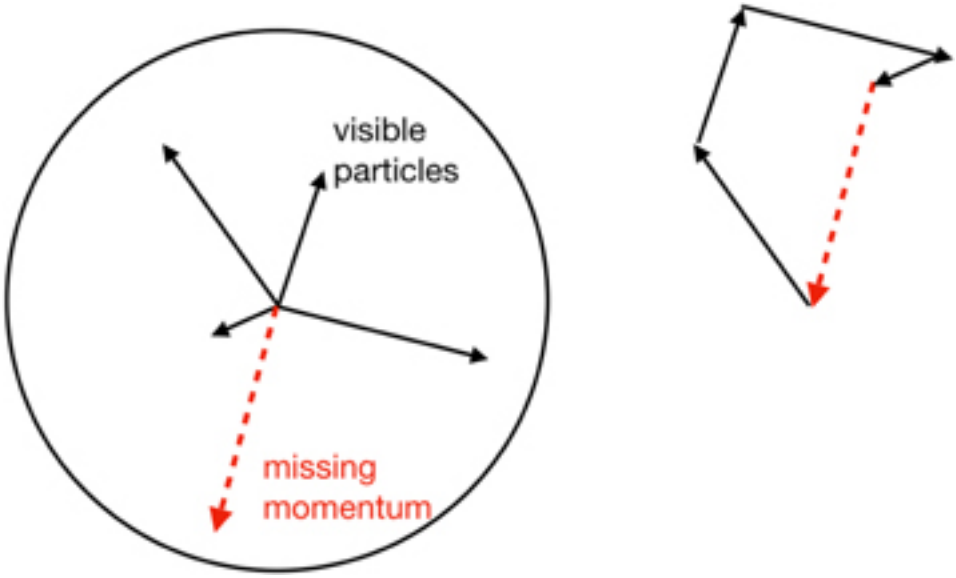


Figure 4 – Diagram showing how it is possible to infer the presence of particles based on 'missing' momentum or energy in the transverse plane. Conservation of momentum and energy laws stipulate the momentum and energy must remain the same before and after a collision [10]

Data

The data used was the Monte-Carlo (MC) set released by the ATLAS collaboration. The data was simulated, using pseudo-random number generators to closely emulate the proton-proton collisions. The simulated data is not as complex as real data from the detector, as only the dominant physical processes are accounted for. Simulated data had the advantage of being able to test whether the analysis techniques were effective for the theoretically predicted results, allowing refinements to be made to the analysis method. [5]

The MC data consisted of a collection of information describing the evolution of proton-proton collisions. The background was ttbar production, which had a very similar decay process to the signal, making differentiating between signal and ttbar events problematic. Picking a

background process similar to the signal was important to ensure the analysis methods were able to differentiate between a distribution similar to the signal, and the actual signal.

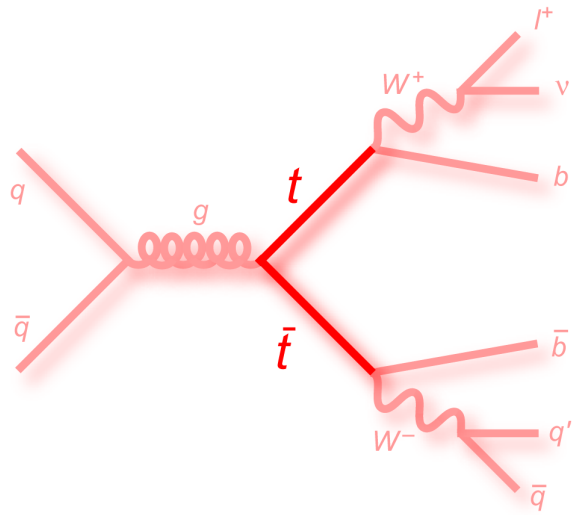


Figure 5 - The background production process (ttbar production), which mimics the signal in data [9]

The MC sample was rescaled to account for the cross-section and integrated luminosity of the ATLAS detector. The amount of data is quantified by the integrated luminosity:

$$L = \int Ldt \quad (1)$$

Where the L within the integral is the instantaneous luminosity depending on the characteristics of the proton beams.

The integrated luminosity used to rescale the data was $150fb^{-1}$, as this was the value used for analysis optimisations at ATLAS [14]. The number of events for a given process expected in a dataset with integrated luminosity L is:

$$N = L \cdot \sigma \quad (2)$$

Where σ is the production cross section (rate) of a given process – which is considerably higher for the ttbar than for the signal.

Object Definitions

A brief description of the objects and their definitions is given:

- Leptons: electrons and muons are included in this analysis. Their minimum momentum is 27 GeV as single lepton triggers are used, and a 25 GeV threshold should be passed for the events to be recorded. Also $|\eta| < 4.8$, where $|\eta|$ is the pseudorapidity [1]
- Jets: A tight cone of hadrons formed when quarks and anti-quarks combine. Quarks cannot be detected directly in particle detectors – their existence is inferred from these resultant jets.
- B-jets: A jet formed by two bottom quarks decaying (one quark and one antiquark). The radius of the jets is R=0.4, where R is defined in terms of ϕ and η
- Missing Transverse Momentum (MET): Momentum that is not detected by the detector but can be inferred due to conservation of momentum laws. Particles carrying this momentum interact very weakly with the electromagnetic and strong forces, so are difficult to detect. Its modulus is E_T^{miss}, the missing transverse energy.

Kinematics and variables

Aside from the MET, other variables were considered. These variables were ones collected by the ATLAS detector when running experiments, which allow for detailed analysis and reconstruction of the collisions. The variables used are given below with explanations:

- Transverse mass (m_T), which peaks around the W boson mass for events with W bosons. The formula for the transverse mass is:

$$m_T = \sqrt{2p^l{}_T MET\left[1 - \cos\left(\Delta\phi_{l,MET}\right)\right]} \quad (3)$$

  Where $\left|\overrightarrow{p_{T,l}}\right|$ is the momentum of the lepton in the event, $\left|\overrightarrow{p_{T,\nu}}\right|$ is the momentum of the neutrino, and $\Delta\phi_{l,MET}$ is the angle between the lepton and the neutrino's direction vector.

- Contransverse mass (m_{CT}), which takes into account the $b\bar{b}$ system:

$$m_{CT} = \sqrt{2P^{b1}{}_T P^{b2}{}_T(1 + cos\Delta\phi)} \quad (4)$$

  Where $\Delta\phi$ is the azimuthal angle between the two jets.

- m_{bb} Is the invariant mass of the two b-jets, which is a useful differentiator, as it peaks around the mass of the Higgs boson if it is present (approx. 125GeV)

Other variables were included in the data, which refer to the magnitude of momentum or the angle of various particles produced in the collision. [1]

Data plots

[1] Pseudorapidity is defined as $\eta = -\ln[\tan(\theta_{cm}/2)]$ where θ_{cm} is the centre of mass angle in the x-z plane, not to be confused with the azimuthal angle ϕ, which is the scattering angle in the x-y plane

The distributions of both the signal and background were plotted to see how they differed. The graphs shown are normalised, so that the signal and background both have an integral of 1. Plotting in this way was important to identify good methods of differentiating between signal and background.

There were some preselections applied to the data:

- Missing transverse momentum (MET) > 100 GeV, meaning any events included are significant and not the result of minor anomalies in the detector's results.
- 2 b-jets with transverse momentum (p_T) > 30 GeV, which is optimal threshold to have high efficiency in the reconstruction of the b-jets.
- 1 lepton with transverse momentum (p_T) >27 GeV, which is again due to the trigger in the detector
- Invariant mass of the two b-jets (m_{bb}) between 50 and 200 GeV, with 3 jets maximum. This is due to the fact that the signal has 2 b-jets as the Higgs boson decays, with potentially another jet emitted by an initial parton. The background (ttbar) normally has at least 4 jets, so this also reduces the number of background events.

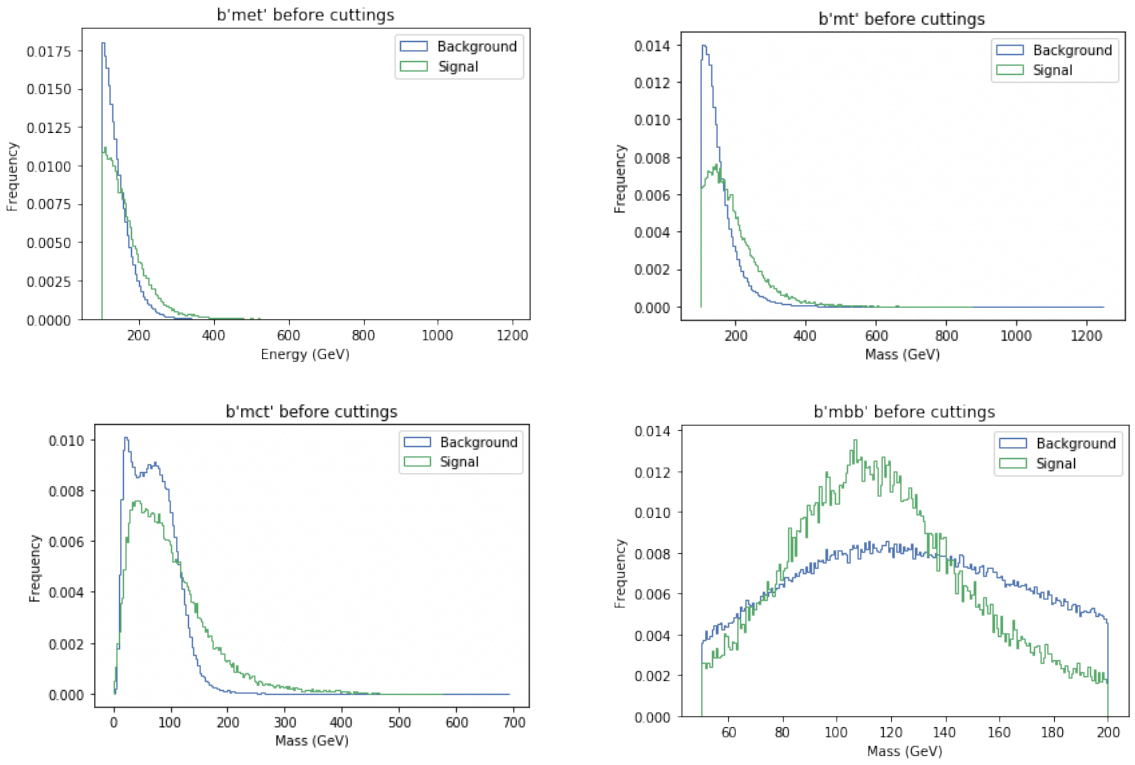


Figure 6 – m_T, MET, m_{bb} and m_{CT} of the low mass sample. Both the signal and background curves are normalised so each has an integral of 1. Frequencies are relative frequencies, not absolute.

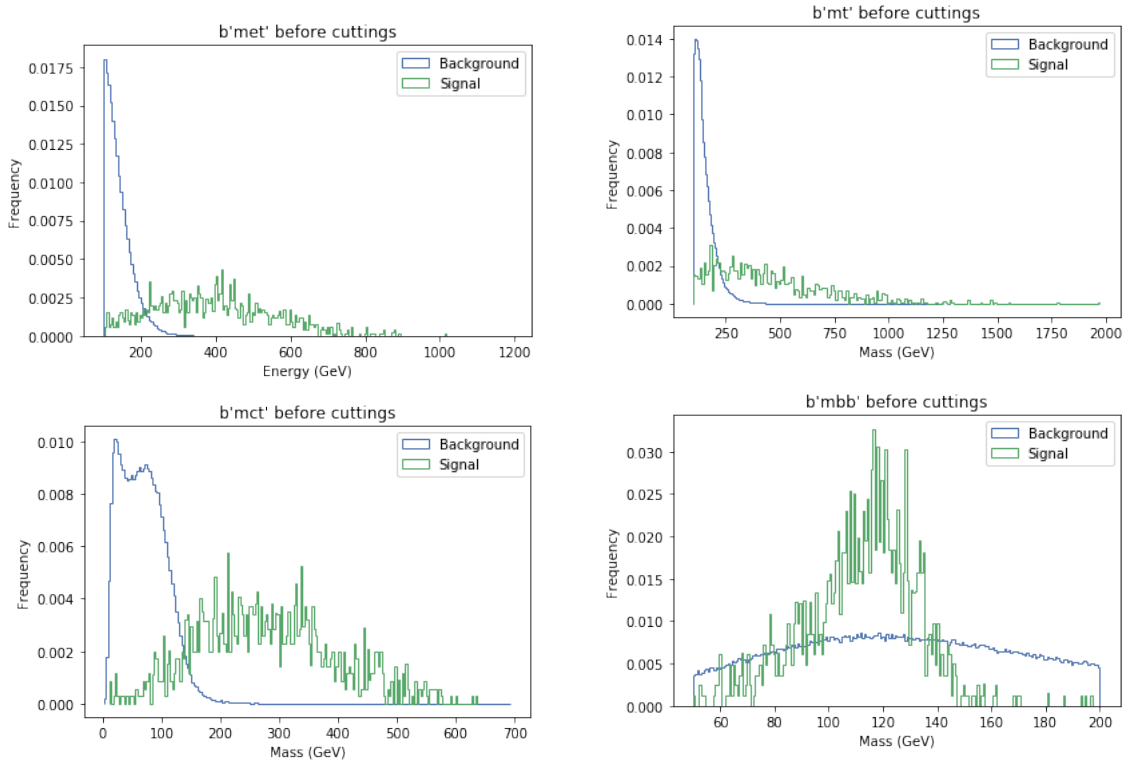


Figure 7 - m_T, MET, m_{bb} and m_{CT} of the high mass sample. Again, both signal and background curves are normalised so each has an integral of 1. Frequencies are relative frequencies, not absolute.

In the low mass sample, clearly the shape of the distributions are similar for both signal and background are very similar, however the signal has a higher proportion of events at higher energies for the m_T, MET and m_{CT}. Rejecting events at low values of these three variables would be effective. The m_{bb} remains centred around 125 GeV – the mass of the Higgs boson which the b-jets decay from.

For the high mass sample, the majority of the signal events are at far higher energies, making its distribution quite different to that of the background. However, the shape of the m_{bb} curve remains centred around the Higgs boson mass, as the two b-jets will have the same invariant mass regardless of the mass of the SUSY particles. Including events only at higher energies is likely to be more effective for this sample for the MET, m_T and m_{CT}.

Cut-and-count analysis

The analysis technique initially used is called a cut-and-count analysis, whereby a cutting is made at a certain energy, and only events with a higher energy are included in calculations. The aim was to increase the ratio of signal events to background, which was measured by a figure of merit called the significance:

$$Significance = \frac{N_s}{\sqrt{N_b + \Delta N_b^2}} \quad (5)$$

The uncertainty was taken to be 30%, a value slightly larger than found in the previous study, which was 15-20%. However, the uncertainty on other models such as the 0lbb (no leptons and 2 b-jets) was 50%, so it is not unusual to encounter large uncertainties, and thus a conservative estimate was made. The significance equation takes into account the uncertainty on the background events, to prevent the significance becoming inflated by an unusually low number of background events.

Cuts were made on the four main variables – m_T, m_{CT}, MET and m_{bb}. Cuts could have been made on the other kinematic variables (such as angle sizes and other masses or momentums), however this would overcomplicate the process, and too many events would be excluded. Also, these variables were strongly correlated with the other kinematics, so cutting on one of these four variables would be a cutting on the other kinematics. To start, a script was written which calculated the significance found when increasing the value at which a variable was cut. This helped determine which cutting values had the greatest effect on the significance.

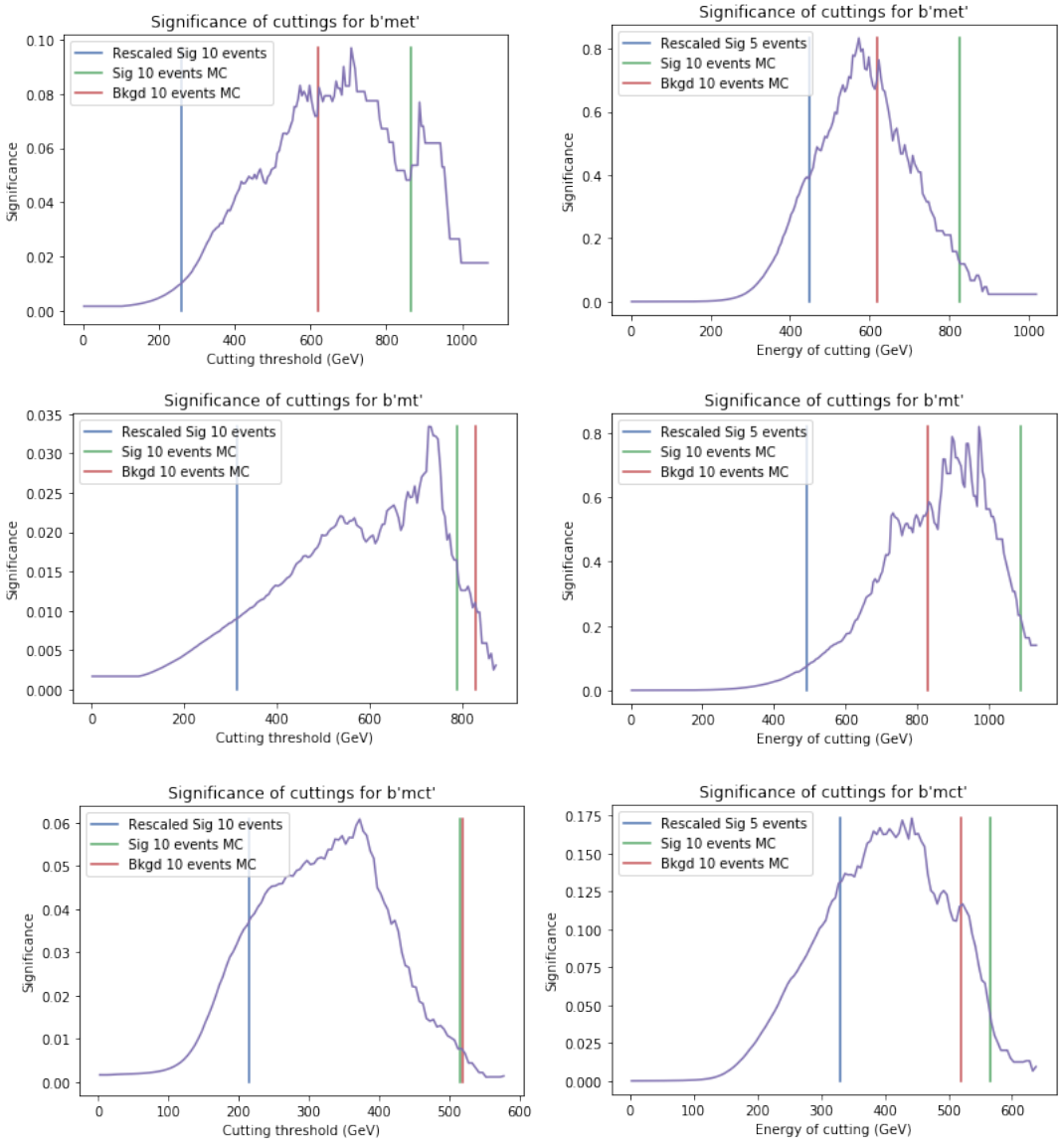


Figure 8 – Graphs showing the significance when cuttings are made at different energies on the low mass sample (left) and high mass sample (right). If the cutting threshold is 200 GeV for the MET, this means all events with MET > 200 GeV are included in the significance calculation (equation 5). The blue vertical line indicates that a cutting made at this point has few than 10 signal events present after rescaling. The green and red vertical lines indicate that fewer than 10 MC events were present (before rescaling) for signal and background respectively. See below for explanation why this is important.

Although the significance may appear extremely low, with no hope of obtaining meaningful results, it is important to remember that the graphs in figures 6 and 7 were normalised. See the table below for information on the number of signal and background events. Clearly, the signal was dwarfed by the number of background events, and reaching even a significance of 1 would be a good result.

	Low mass sample	High mass sample
No. of signal events	157	15
No. of background events	313741	313741
Initial significance	0.00167	0.000157

As well as the low significance, there was another issue. As soon as the significance started to rise, the number of signal events started to fall drastically. As can be seen, the highest levels of significance are very far past the vertical lines, so likely have very low numbers of signal. If fewer than 10 events are present after rescaling (or 5 on the high mass sample), then cutting on four variables simultaneously was likely to eliminate all signal events. This is because the variables were fairly correlated, and so cutting on one variable had an effect on the other. If cutting on just one variable rejected a lot of signal, cutting on four at once could reject all the signal.

Also, since the data was rescaled, it was important to not reduce the number of events in the MC sample below 10 for either signal or background. Since the MC sample was probabilistic, it was not guaranteed that any of these events would actually occur. For example, if in the MC sample there were 2 signal events present after cutting, this may then be rescaled to 50 events after luminosity and cross-section calculations. Despite this looking quite positive in terms of the number of rescaled events, it is entirely feasibly that the 2 MC events would not occur, and thus, neither would the 50 rescaled events.

As expected, the significance of the high mass sample reached higher levels, despite the number of signal events being over 10 times less, due to the distribution of events being concentrated at higher energy levels than the background. The script was not used for the m_{bb}, as two cuts were needed here, due to the signal events being concentrated around the 125 GeV mark.

However, these significance results were only for cuttings on one variable at a time. For a more complete analysis, potentially with better results, cuts were made on the four variables simultaneously. To avoid eliminating all signal events, cuttings were made at a lower energy value than the blue line. If this were not the case, cutting on all 4 variables at once would exclude too many signal events. Thus, cuttings were made at the following values for the low mass sample:

- MET > 200 GeV

- $m_T > 280$ GeV
- $m_{CT} > 200$ GeV
- $50\text{GeV} < m_{bb} < 200$ GeV

And for the high mass sample:

- MET > 240 GeV
- $m_T > 240$ GeV
- $m_{CT} > 180$ GeV
- $100\text{GeV} < m_{bb} < 140$ GeV

The output is shown below:

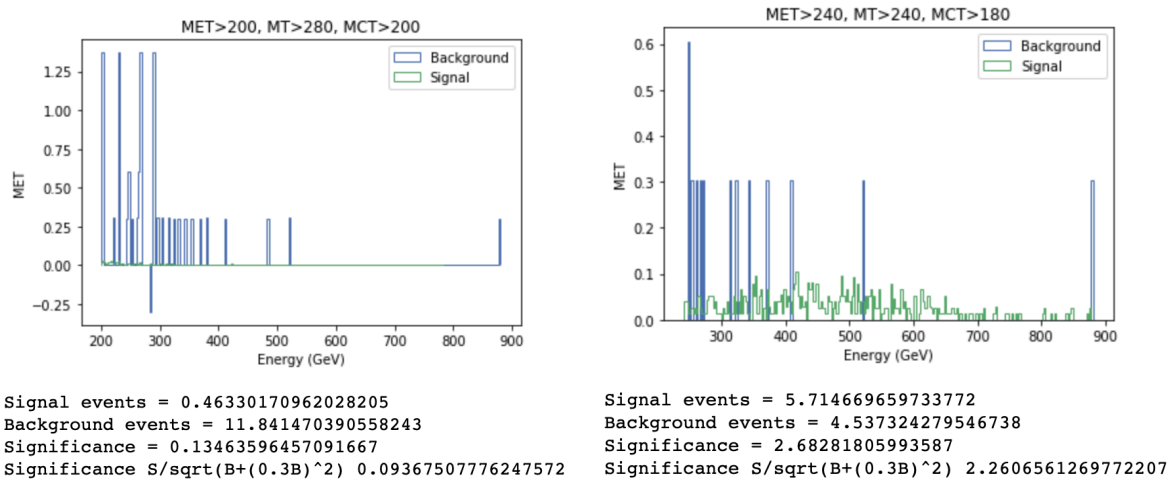


Figure 9 – Graphs showing the MET of signal and background events once selections have been made for the low mass sample (left) and high mass sample (right). The first significance value does not include the uncertainty in the calculation, and so the second significance value holds more merit.

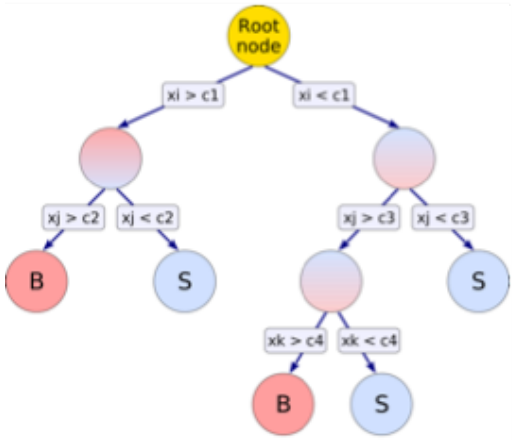
The table below shows the results before and after the cuts were applied. As can be seen, for the low mass sample, nearly all of the signal has been rejected, as well as nearly all of the background. Although the significance is higher, it is only higher by nearly 2 orders of magnitude, and is still far below 1 – which would indicate around 1 signal event for every background event. For the high mass sample, the results are far better, as expected. Only around 1/3 of the signal has been rejected, whilst the vast majority of the background has been rejected. This results in a significance of 2.261, an increase of four orders of magnitude over the cut-and-count.

	Low mass sample	High mass sample
No. of signal events	157→ 0.463	15 → 5.715
No. of background events	313747 → 11.841	313747 → 4.537
Significance	0.00167 → 0.0937	0.000157 → 2.261

Figure 10 – Table showing the difference in signal events, background events, and significance before and after cuttings/selections were made. Both the high mass sample and low mass sample are shown.

Machine learning analysis

After mixed results from the cut-and-count method, a new approach was tried – using machine learning (ML). Two machines were trained. The first was a boosted decision tree (BDT). A BDT consists of a set of nodes, or questions, with each question only having two answers. The next question or node depends on the answer of the previous question. The final answer, called a leaf, is reached when there are no more questions and the end of the nodes has been reached. In this way, a BDT can classify data, and is a good way to perform multivariate analysis. [3]
Figure 11 – the classification process of a BDT. [3]



The other type of network trained was a deep neural network (DNN). A DNN assigns different weights to input variables, thereby giving each variable a different importance or significance. These weights change to what extent the input variable has an effect on the next set of neurons. In the final layer, the product of the input variables and the weights assigned by the neurons classifies the data accordingly. [6]

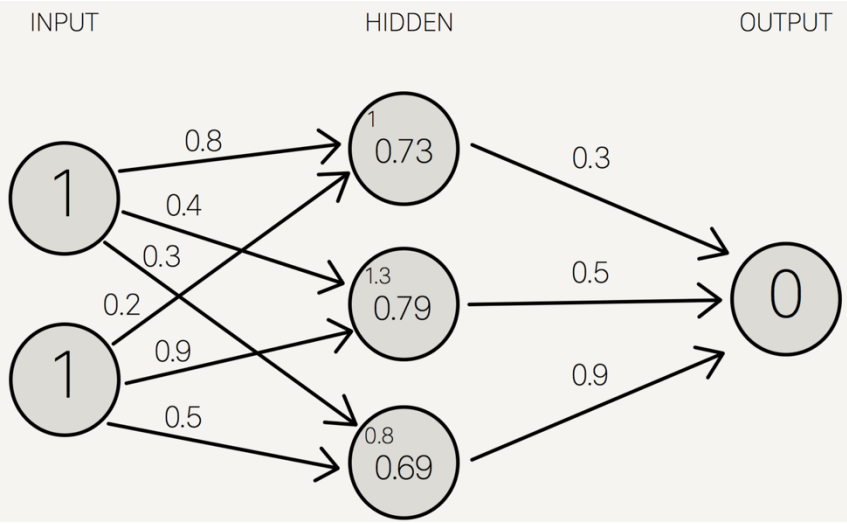


Figure 12 – Diagram detailing how a DNN is structured, with different weights assigned to input variables to classify the data. [6]

The two networks were first trained once on each signal and on signal and background datasets. The training told the machines how each event was classified. The machines then used this data to develop and train its classification systems in order to gain the highest accuracy. The accuracy, or how well the machine has modelled the data is given by the loss function, with a number closer to zero being better. The machines go through this process repeatedly (which is called an "epoch"). After each epoch, the weights or decision making parameters are updated in the networks. Also, it calculates the loss function of its new classification rules. To prevent overtraining, if the loss function had not decreased after 5 epochs, the training was stopped. Another way to prevent overtraining was by limiting the learning rate (LR). The LR quantifies how quickly a network replaces old rules or beliefs for new ones – or put simply, how quickly it 'learns' how to analyse the data. A high LR allows the network to be tuned very well, but if it is too high, it may acquire bad classification rules quickly. Take the example of a DNN trained to recognise pictures of dogs. If the first few items it is trained to recognise are black dogs, then the next 50 are white dogs, the network may 'forget' that the colour of a dog is not important when deciding its species, and quickly decide its colour is the most important trait. A low LR of 0.0001 was chosen, to prevent the network becoming overtrained.

One way of seeing the results of the classification was to view a graph of the background rejection against the signal efficiency. A straight line from the maxima of each axis would indicate 50/50 classification – random guessing. The more curved towards the top the curve is (a high background rejection for high signal efficiency) the more accurately the machine can be said to be classifying the data. However, this could also be the result of the networks being overtrained on one set of data, so care had to be taken not to increase the learning rate and other relevant variables to increase the classification accuracy and inadvertently overtune the machines.

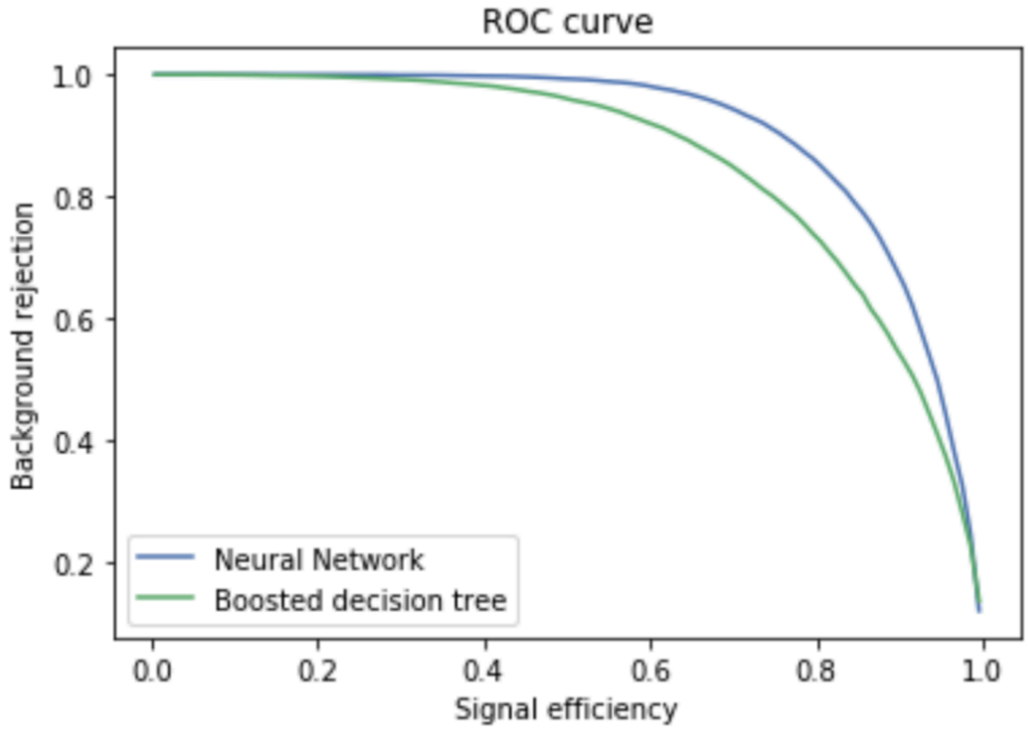


Figure 13 – The background rejection against signal efficiency for the low mass sample, calculated during the validation part of the training process.

The variables (such as learning rate) were altered to train the machines with many different settings, and graphs of these variable distributions can be seen in the appendix. For the data analysed, the following settings were chosen:

- DNN learning rate: 0.0001.
- DNN early stopping: 5. This is the number of times the machine will continue training (or epochs) when the loss function remains the same and no improvement is made. A low value prevents overtraining, whilst a high value can allow for a very tuned machine.
- DNN layers: 2. This is the number of hidden layers of neurons in between the input and output layer.
- DNN hidden nodes: 7. The number of nodes per hidden layer.
- BDT trees: 1000. The number of trees in the BDT.

Also important to note is the variables inputted into the machines. Previously, the cut-and-count only made selections based on the four most important variables. However, multivariate analysis using computers can uncover intricate relationships between sets of data that humans simply would be unable to see. Further, the machines can perform a high number of calculations very quickly, so increasing the number of variables would not overcomplicate the analysis. To illustrate this point, see below the background rejection against signal efficiency when only the MET, m_T, m_{CT} and m_{bb} are inputted as variables. The more curved towards the top right the line is, the better it has classified the data – as the aim is to reject no signal whilst rejecting all background. Clearly, increasing the number of input parameters does not hinder, but actually aids the machines in classifying.

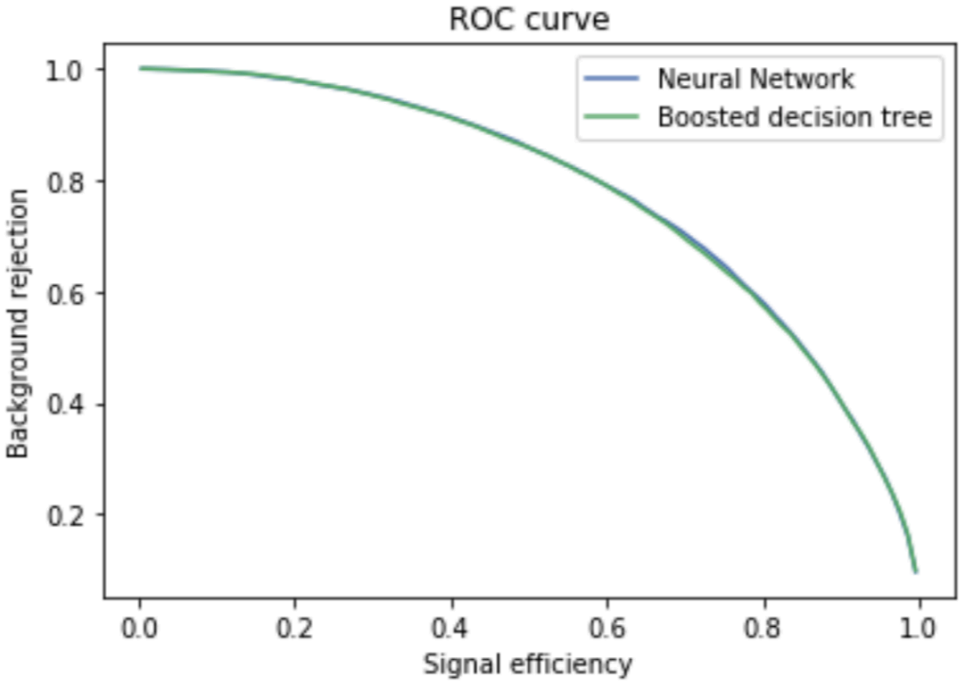


Figure 14 – background rejection against signal efficiency for the low mass sample when only 4 variables are considered - MET, m_T, m_{CT} and m_{bb}.

Once the networks had been trained, they were instructed to analyse the two data sets, and, based on its classification criteria, classify the events as either signal or background.

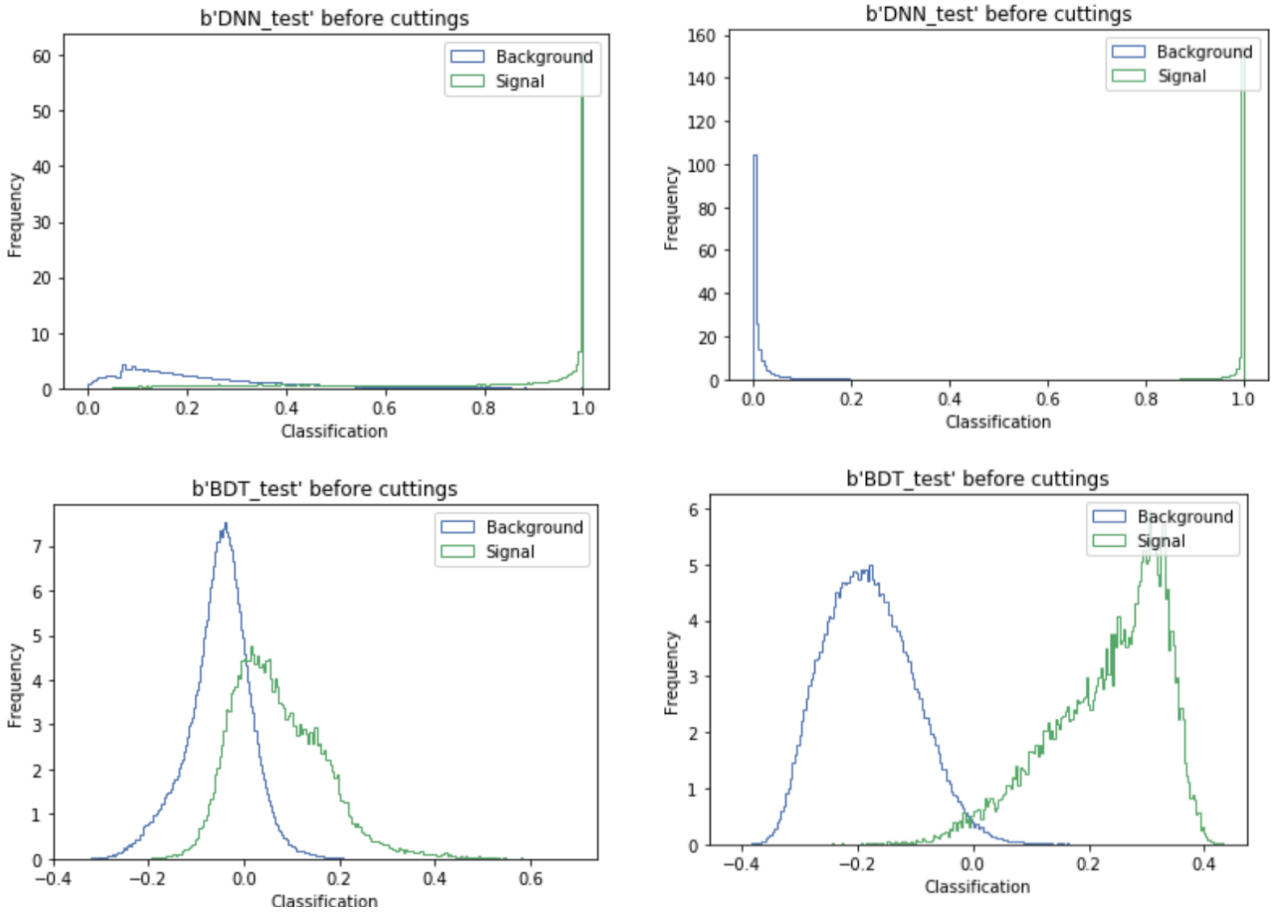


Figure 15 – the classification of the signal and background as done by the DNN and BDT. Low mass sample is left, high mass sample is right. Both the signal and background curves have been normalised to give an integral of 1. Frequencies are relative, not absolute. For the DNN, a classification of 1 means a high likelihood of an event being signal, whilst 0 is a high likelihood of it being background. For the BDT, a classification of 1 means a high likelihood of an event being signal, whilst -1 is a high likelihood of it being background.

The results of the low mass sample look quite promising for the DNN. A very high proportion of signal events have been correctly classified as signal, and most of the background events are indicated as being likely to be background. However, the BDT has struggled to classify the signal and background with as much certainty, with many events being close to 0, meaning the BDT is unsure if the event is signal or not.

The high mass sample has better results, as expected due to the distribution of the signal being less similar to the background compared to the low mass sample. Again, a very high proportion of signal events are classified correctly as signal, but a higher proportion of the background is closer to a classification of 0 for the high mass – indicating a higher level of confidence in the classification. Again, the BDT seems to have poorer results, but the mean of the signal and background are further away from 0 compared to the low mass sample.

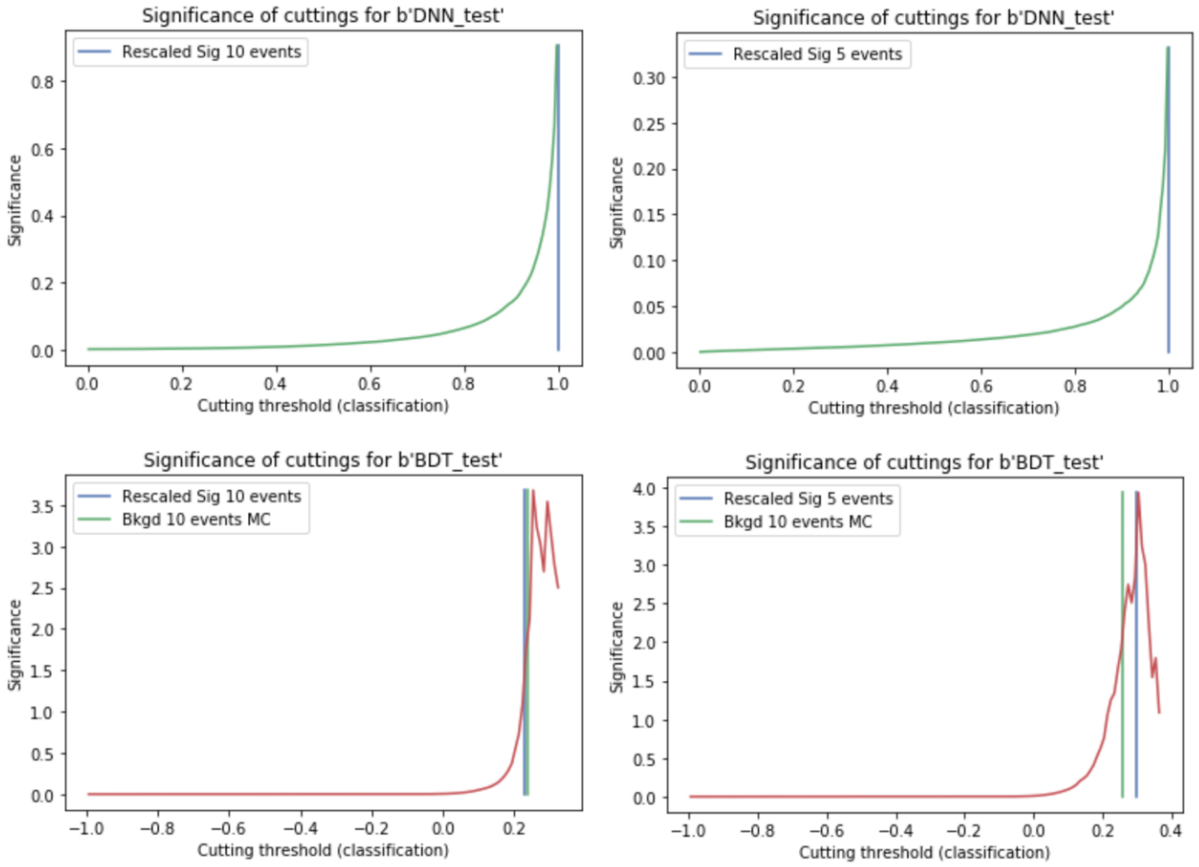


Figure 16 – The significance at different cutting thresholds for the low mass sample (left) and high mass sample (right). Graphs of both the DNN (top) and BDT (bottom) are shown.

For the low mass sample, the DNN has far better than the cut-and-count in classifying events as signal or background, and the number of MC events never goes below 10. The BDT has also reached a high significance, but the number of MC background events is below 10 at the highest significance levels. Due to the probabilistic method of how the MC sample is generated, these events may not occur in a real experiment, or may be much higher. Therefore, the number of background events (regardless of how many are present after rescaling) is unstable. Although the high significance is good, there is a risk that the result itself is unstable. On the other hand, it could be that very little background would actually be produced in reality. Thus, this result should be seen as promising, but taken with caution.

For the high mass sample, the results are far worse for the DNN. The significance achieved is very low, in fact even lower than the cut-and-count. The BDT results are similar to the low mass sample – a high significance is reached, but again the number of background MC events falls below 10. Like for the low mass sample, this is a good result, but it is unstable.

The table below summarises the results for the two machines. The values given are those where the number of MC signal and background events is more than 10, as these hold the most weight. Much higher levels of significance were found for both the cut-and-count and machine learning, however these had far below 10 signal or background events in the MC sample.

DNN	Low mass sample	High mass sample
Highest significance	0.906	0.331
No. of signal events	47.7	11.8
No. of background events	169.9	113.3

BDT	Low mass sample	High mass sample
Highest significance	1.76	1.91
No. of signal events	9.4	7.8
No. of background events	13.1	9.2

Figures 17 (top) and 18 (bottom) – Tables showing the results obtained by machine learning techniques. These were calculated by making a cutting on the classification of events by the machines. These results give the highest significance that is statistically significant – i.e. with at least 10 MC signal and background events present.

Discussion

For the cut-and-count analysis, quite clearly the high mass sample had better results, with much higher levels of significance that for the low mass sample. This was due to the considerable difference in distributions between the two samples. The high mass sample had much higher energy MET, m_{CT} and m_T values than the ttbar, allowing easy differentiation between the two sets. Even plotting the MET, m_{CT}, m_T and m_{bb} distributions allows for easy differentiation by eye.

The low mass sample had bad results for the cut-and-count, largely due to the similarity between signal and ttbar distributions. For sets of data that were very similar, cut-and-count was not effective, and this was exacerbated by the sheer volume of background events in comparison to the signal.

For the ML analysis, the low mass sample was a success, having far higher significance levels when analysed by both the DNN and BDT. The BDT was far superior in its analysis here, obtaining nearly double the significance of the DNN. However, since the original number of signal and background events were 157 and 313737 respectively, the DNN has managed to retain a lot more of the signal, whilst still rejecting most of the background. The BDT seems to obtain a higher signal to background ratio, but also rejects a lot more of the signal.

For the high mass sample, the results were quite poor for both the DNN and the BDT, but especially for the DNN. The significance for both machines is actually lower than for the cut-and-count. Again, the DNN manages to retain more signal, but is worse at rejecting background, and so the BDT obtains a higher significance.

The high mass having poorer results for the ML than the cut-and-count was interesting, since common sense would dictate that like humans, the machines should find it easier to differentiate between very different data sets. There are several possible reasons for this

First, the number of MC events the machines had to train was different for the low mass and high mass samples. In the low mass sample, there were 42127 MC signal events, compared to

22993 for the high mass – a difference of nearly half. Since training is imperative for these machines, this could have had an impact on how well it could classify.

Secondly, for the high mass sample, due to technical difficulties, there was no m_{CT} variable available to train the machines. Although there were nine other variables used to train, arguably this is one of the more important variables, and this could have had an impact.

Finally, the requirement of needing at least 10 MC signal and background events to be present. The BDT in particular managed to obtain a significance of 3.93 at its maximum (see figure 16), but this only contained 4 MC background events and 6985 MC signal events. The number of MC background events being so low means in reality, this number could be much higher, which would translate to more real background events after rescaling. This means the result has a high potential error. However, this could also be simply because there would be very few background events present after such a cutting. This is perhaps an area for further research, as such results although not statistically significant at this point in time, could be promising.

Conclusion

As a result of this study, both traditional cut-and-count analysis techniques and machine learning have been applied to data. The data was generated to probabilistically emulate proton-proton collisions, and this was rescaled to account for the luminosity of the ATLAS detector ($150fb^{-1}$) and the cross sections of the collisions. Two hypotheses of supersymmetric particle production were tested, along with the ttbar background, which is very similar to the signal, to ascertain whether the analysis techniques used could differentiate between signal and background events. The machine learning approach was very successful for the low mass sample, obtaining a far higher significance than for the cut-and-count. The high mass sample had far better results for the cut-and-count, however there were variables that could have affected the machine learning aspect for this sample, such as the issues relating to the MC statistics. In short, machine learning has the potential to differentiate supersymmetric particle production and other particle collisions, and could be key to obtaining evidence to support the lightest supersymmetric particle being dark matter.

References

[1] ATLAS Collaboration (2018). *Search for chargino and neutralino production in final states with a Higgs boson and missing transverse momentum at √ s = 13 TeV with the ATLAS detector*. [online] Available at: https://arxiv.org/pdf/1812.09432.pdf [Accessed 1 May 2019].

[2] Bennett, J., Donahue, M., Schneider, N. and Voit, M. (2014). *The essential cosmic perspective*. 7th ed. Pearson.

[3] Boser, C., Fink, S. and Rocker, S. (2019). *Introduction to Boosted Decision Trees - A multivariate approach to classification problems*. [online] Indico.scc.kit.edu. Available at: https://indico.scc.kit.edu/event/48/contributions/3410/attachments/1690/2312/BDT_KSETA_Freudenstadt.pdf [Accessed 3 May 2019].

[4] CERN (2019). *ATLAS*. [online] CERN. Available at: http://www.home.cern/science/experiments/atlas [Accessed 1 May 2019].

[5] CERN (2019). *Data and simulated data*. [online] Opendata.atlas.cern. Available at: http://opendata.atlas.cern/books/current/openatlasdatatools/_book/data_and_simulated_data.html [Accessed 1 May 2019].

[6] Cosmosmagazine.com. (2017). *What is Deep Learning and how does it work? | Cosmos*. [online] Available at: https://cosmosmagazine.com/technology/what-is-deep-learning-and-how-does-it-work [Accessed 3 May 2019].

[7] Griest, K. (1995). The nature of the dark matter. *arXiv*. [online] Available at: https://arxiv.org/pdf/astro-ph/9510089.pdf [Accessed 1 May 2019].

[8] Gunn, J., Lee, B., Lerche, I., Schramm, D. and Steigman, G. (1978). Some astrophysical consequences of the existence of a heavy stable neutral lepton. *The Astrophysical Journal*, 223, p.1015.

[9] Heinson, A. (2019). *ttbar production*. [online] Www-d0.fnal.gov. Available at: https://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/feynman_diagrams/feynman_ttbar_ljets_longt.png [Accessed 1 May 2019].

[10] Penning, B. (2018). The pursuit of dark matter at colliders—an overview. *Journal of Physics G: Nuclear and Particle Physics*, 45(6), p.063001.

[11] Quanta Magazine. (2012). *Physicists Debate Future of Supersymmetry*. [online] Available at: https://www.quantamagazine.org/physicists-debate-future-of-supersymmetry-20121120/ [Accessed 8 May 2019].

[12] Rubin, V. and Ford, W. (1970). Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *The Astrophysical Journal*, 159, p.379.

[13] Seigar, M. (2016). *Dark Matter in the Universe*. Morgan & Claypool Publishers, pp. 3.0-5.0.

[14] Twiki.cern.ch. (2019). *LuminosityPublicResultsRun2*. [online] Available at: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 [Accessed 26 Apr. 2019].

Appendix

The code and output used to analyse the data is available at:
https://github.com/James-Tipping/University-thesis

Clicking on the above link will open a browser window with three files available:

- For a main summary of the cut-and-count and ML analysis, or to see more examples of plots included in this document, click on Cut-and-count and ML analysis summary.ipynb

- For plots of all of the kinematics and variables for the low mass sample, click on Low mass variable plots and correlation plots.ipynb

- To see how changing parameters of the DNN and BDT during training altered their effectiveness in classifying data as signal or background, click on Classification during training plots - low mass sample.ipynb