

## FINAL YEAR PROJECT REPORT

NAME:	Corey Swinbank
DEGREE COURSE:	BSc Physics
PROJECT TITLE:	Boosting dark matter searches using machine learning
YEAR OF SUBMISSION:	2023
SUPERVISOR:	Henning Flaecher
NUMBER OF WORDS:	7154



## **Declaration**

This research project was supervised by Henning Flaecher and completed by myself, Ammon Radclyffe and Telsia Brake. All simulated LHC event data was provided by Ph.D. student Maciek Glowacki. The Bristol DICE cluster was used to carry out all computational work. Extensive use of Tensorflow, scikit-learn, pandas and NumPy were made throughout the project. Ideas for the general direction of the project were often set by Henning, but the implementation of the ideas was independent. Luke Kreczko of the Bristol Dice cluster provided the initial code to access the data, however beyond this point all code was written independently and can be found on my GitHub repository at [github.com/CoreySwinbank](https://github.com/CoreySwinbank). All results presented in this report are my own.

## **Acknowledgements**

I would like to first thank Henning, Ammon and Telsia for their stimulating conversation during our weekly meetings. A special thanks to Henning for providing all physics insight throughout this project. I would also like to thank Maciek for providing assistance on computational issues, and Luke for assisting me in setting up my coding environment and opening the dataframes to begin work on the project.

# Boosting Dark Matter Searches using Machine Learning

C. Swinbank

School of Physics, University of Bristol

Date: April 19, 2023

## Abstract

This report investigates the use of neural networks in the event classification of top-antitop quark Higgs production at the LHC, with the Higgs decaying invisibly to dark matter. The research is carried out to aid the search for beyond the standard model physics in Higgs to invisible decays. Feedforward and recurrent neural networks are implemented, both individually and combined. The combined model displayed the best performance, achieving a ROC area under curve of  $0.7945 \pm 0.0008$  and a maximum significance of  $(10.4 \pm 0.4)\sigma$ . This corresponds to  $210 \pm 10$  signal and  $190 \pm 10$  background events classified as Higgs to invisible decays, normalised to  $300 \text{ fb}^{-1}$  of data to simulate run 3 at the LHC.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Background</b>	<b>4</b>
2.1	Higgs Production Modes . . . . .	4
2.2	Machine Learning . . . . .	5
<b>3</b>	<b>Experimental Methods</b>	<b>5</b>
3.1	CMS Detector . . . . .	5
3.2	Data Simulation . . . . .	6
3.3	Event Selection . . . . .	6
3.4	Deep Learning . . . . .	7
3.4.1	Feedforward Neural Networks . . . . .	7
3.4.2	Recurrent Neural Networks . . . . .	8
3.5	Types of Classifier . . . . .	8
3.6	Class Imbalance and Cross-section Weights . . . . .	8
3.7	Performance Metrics . . . . .	8
3.8	Hyperparameters . . . . .	9
<b>4</b>	<b>Results</b>	<b>9</b>
4.1	Feed-forward Neural Network . . . . .	10
4.2	Recurrent Neural Network . . . . .	11
4.3	Combined Model . . . . .	11
4.4	Multiclass Classifier . . . . .	12
4.5	Model Stability . . . . .	13
<b>5</b>	<b>Discussion</b>	<b>14</b>
5.1	Future Work . . . . .	15
<b>6</b>	<b>Conclusions</b>	<b>15</b>
	<b>Appendices</b>	<b>16</b>

# 1 Introduction

Since the inception of the standard model (SM) in the early 1970s, it has been the most rigorously tested and successful theory in particle physics. Despite this, it still has many shortcomings, showing the model is incomplete. Key shortcomings include; the SM does not include gravity, it fails to explain the baryon asymmetry observed in the Universe, and it does not offer a dark matter (DM) candidate [1]. These shortcomings motivate searches for physics beyond the SM (BSM).

Cosmological observations throughout the 20th century provided evidence for the need of a new type of non-baryonic matter [2, 3]. The latest measurements of the cosmic microwave background (CMB) from the Planck Collaboration found DM accounts for 84% of matter in the Universe, motivating the need to understand it further [4]. Overall, the observations suggest that a DM candidate should be massive, stable, electrically neutral and only interact weakly with SM particles [5]. Many theoretical frameworks that provide a DM candidate of this nature. One set of candidates are weakly interacting massive particles (WIMPs). WIMPs appear in many theories, such as the theory of supersymmetry (SUSY), which postulates a new symmetry between fermions and bosons. Many SUSY models predict that the lightest supersymmetric particle is a WIMP candidate for DM [6]. Several models in these theoretical frameworks, known as Higgs portal models, predict that DM interacts with SM particles via the Higgs Boson [7].

The Higgs boson was discovered in proton-proton (pp) collisions at the Large Hadron Collider (LHC) at a mass of 125 GeV [8]. The Higgs is responsible for giving mass to fundamental particles, which could enable it to be a portal between SM particles and DM. If this is the case, then a pair of DM particles could be produced from a Higgs decay in pp collisions, assuming it is kinematically favourable, i.e.  $m_{DM} < m_H/2$  [9]. Due to the weakly interacting nature of DM, it does not leave a signature in a particle detector, therefore its existence is measured indirectly as missing transverse momentum,  $p_T^{miss}$ . A particle detector, such as CMS or ATLAS, will sum the transverse momenta of all decay products in an event, and compare this with the initial transverse momenta to test whether any momenta has left the detector in the event [10, 11]. Due to the undetectable nature of a Higgs decaying to a pair of DM particles, this is known as a Higgs to ‘invisible’ decay.

In the SM, a Higgs to invisible decay can only occur through  $H \rightarrow ZZ^* \rightarrow 4\nu$ , which has a branching fraction,  $\mathcal{B}(H \rightarrow inv) = 0.1\%$  [12]. A Higgs decay to a pair of DM particles would enhance this branching fraction. If this enhancement was significant, the branching fraction could be observed to exceed the limit set by the standard model. This would be an indication of BSM physics.

At the LHC, the ATLAS and the CMS collaborations are focused on constraining the Higgs to invisible branching fraction. Currently, the most stringent constraint sets  $\mathcal{B}(H \rightarrow inv) = 0.107$  to a 95% confidence level, set by ATLAS [13]. The CMS collaboration has also placed an upper limit of  $\mathcal{B}(H \rightarrow inv) = 0.15$  at a 95% confidence level [14]. These constraints were set using a combination of different Higgs to invisible decay modes. One factor that limits the strictness of this measure-

ment is the ability to distinguish Higgs to invisible processes from their associated SM background processes.

Machine learning (ML) is a novel technique that has recently been implemented to improve branching fraction constraints [15]. It is used across a broad range of tasks, from medicine to physics [16, 17], and has rapidly grown in popularity due to its ability to identify patterns in large sets of data. This makes it a suitable tool to use in particle physics at the LHC [18].

This work was carried out to improve the ability to distinguish Higgs to invisible processes from their associated SM background in top-antitop quark ( $t\bar{t}$ ) events. Machine learning algorithms are used to classify events as signal ( $t\bar{t}H$ ) or background ( $t\bar{t}$ ). The models are trained on Monte Carlo simulated data and results simulate run 3 at the LHC. In the subsequent analysis, a Higgs to invisible branching fraction of 100% is assumed, unless otherwise stated.

## 2 Background

The LHC has completed 2 runs thus far and is currently completing run 3. In these runs, high energy protons collide and detectors measure the information of the decay products produced. The quantity of data stored is measured by integrated luminosity,  $\mathcal{L}$  [19]. The expected number of events,  $N$ , can then be calculated using

$$N = \sigma \int \mathcal{L} dt. \quad (1)$$

where  $\sigma$  represents the cross section for the process [9].

In run 2, the LHC was operating at a centre-of-mass energy of 13 TeV, and collected  $139 \text{ fb}^{-1}$  of data [20]. Run 3 will collect a further  $300 \text{ fb}^{-1}$  at 13.6 TeV, which is particularly significant when considering events with small cross sections, such as Higgs to invisible decays.

In the pp collisions, high energy quarks and gluons are produced. These products fragment and hadronise, producing a shower of particles in a narrow cone. This shower of particles is known as a jet and is detected as a single entity [21].

### 2.1 Higgs Production Modes

At the LHC, there are four main Higgs production modes; gluon fusion (ggH), Vector Boson Fusion (VBF), associated production with a vector boson (VH), and associated production with a pair of top quarks ( $t\bar{t}H$ ) [22]. The cross section of these production modes can be found in table 1, and leading order Feynman diagrams in figure 1.

Table 1: Cross sections of dominant Higgs production modes at centre-of-mass energy of  $\sqrt{s} = 13 \text{ TeV}$  [23].

Production Mode	Cross section of 125 GeV Higgs at $\sqrt{s} = 13 \text{ TeV}$ (pb)
ggH	48.6
VBF	3.78
WH	1.37
ZH	0.88
$t\bar{t}H$	0.51

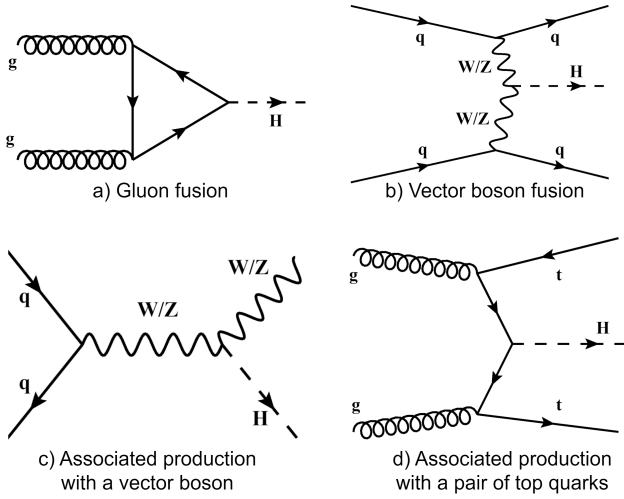


Figure 1: Leading order Feynman diagrams for the four dominant Higgs production modes at the LHC. Adapted from [13, 14].

### Gluon Fusion

Gluon fusion is the most dominant Higgs production mode at the LHC. In this production mode, events with a minimum of one jet and large missing transverse momentum are targeted in searching for BSM physics. Despite having the largest cross section,  $ggH$  is not the most sensitive channel for Higgs detection, due to its irreducible SM background,  $Z \rightarrow 2\nu$  [24].

### Vector Boson Fusion

The most sensitive Higgs production mode at the LHC is VBF. The distinct event topology of VBF makes its signature easier to distinguish from SM processes than other production modes. The distinctive characteristics are two forward jets with large separation on the detector, and the missing transverse momentum from the Higgs to invisible decay. From VBF alone, ATLAS placed an upper limit of  $\mathcal{B}(H \rightarrow inv) = 0.145$  at a 95% confidence level, displaying the sensitivity of this channel [25].

### Associated Vector Boson Production

The VH production mode is challenging to analyse, as it is characterised by one or two leptons and missing transverse momentum. This signature is the same as the irreducible SM background produced by  $ZZ$  and  $WZ$  decays, in which a Z boson decays to two neutrinos, producing real missing energy [26].

### Associated Top Quark Pair Production

The  $t\bar{t}H$  production mode has the lowest cross section of the four main production modes. The dominant SM background in this production mode comes from three channels; the 0-lepton channel (hadronic)  $t\bar{t} \rightarrow WbWb \rightarrow (q\bar{q}b)(q\bar{q}b)$ , the 1-lepton channel (semi-leptonic)  $t\bar{t} \rightarrow WbWb \rightarrow (q\bar{q}b)(l\nu b)$ , and the 2-lepton channel (dileptonic)  $t\bar{t} \rightarrow WbWb \rightarrow (l\nu b)(l\nu b)$ . The hadronic channel is the most common, as the W decays to a quark-antiquark pair 67% of the time, compared to 11% for leptonic decays [27, 28, 29]. Both background and signal events leave only the  $t\bar{t}$  decay signature in the detector, making it challenging to distinguish the signal,  $t\bar{t}H$ , from its associated back-

ground,  $t\bar{t}$ .

The small cross section, along with the difficulty to distinguish this Higgs production mode from its SM background have meant that  $t\bar{t}H$  has been the least analysed of the dominant production modes. The challenge of this production mode is reflected in the upper limit set on  $\mathcal{B}(H \rightarrow inv)$  from this mode by ATLAS. At a 95% confidence level, an upper limit of  $\mathcal{B}(H \rightarrow inv) = 0.38$  was found [30]. This is much greater than the previously referenced upper limit set using the VBF production mode of 0.145.

This is the motivation to analyse the  $t\bar{t}H$  production mode in this research, by creating a machine learning algorithm that can accurately distinguish between signal and background.

## 2.2 Machine Learning

Machine learning (ML) is a novel technique used in data analysis. ML algorithms use mathematical formulae to find complex patterns in data without the need for explicit user programming. This makes ML perfect in tasks where subtle patterns separate different datasets, such as classifying signal and background events in particle collisions. In ML, data is split into three subsets: training data, which the model learns from; validation data, which is used to evaluate the model's training performance; and test data, which is real data used to test the model. [31].

In ML, there are 3 wider types of learning algorithms that can be used; supervised learning, unsupervised learning and reinforcement learning. In this project, only supervised learning is used. Descriptions of unsupervised and reinforcement learning are omitted for brevity, but a full description of all 3 types can be found in [32]. In supervised learning, labels (targets) are provided in training along with the associated input data (features). This enables the model to identify relationships between the targets and features, which generalise to make predictions on unlabelled data. This type of learning is most suitable for event classification at the LHC, as labels are available from event reconstruction.

ML can significantly improve analysis in particle physics. In the VBF production mode, deep learning was used to enhance the search for Higgs to invisible decays. The research used the same data as a CMS search and managed to improve the upper limit on the Higgs to invisible branching fraction by a factor of three [15]. This potential for improvement is the motivation for the use of ML in this project.

## 3 Experimental Methods

### 3.1 CMS Detector

The Compact Muon Solenoid (CMS) detector is a general purpose hermetic detector that is used to search for directly produced new particles at the LHC. It is composed of five broader layers, as seen in figure 2. As particles travel through the different layers of the detector they deposit energy, via ionisation. This leaves tracks in the detector, which allows for the reconstruction of the trajectories of the particles. By measuring the trajectory of the particles in the magnetic field, produced by the

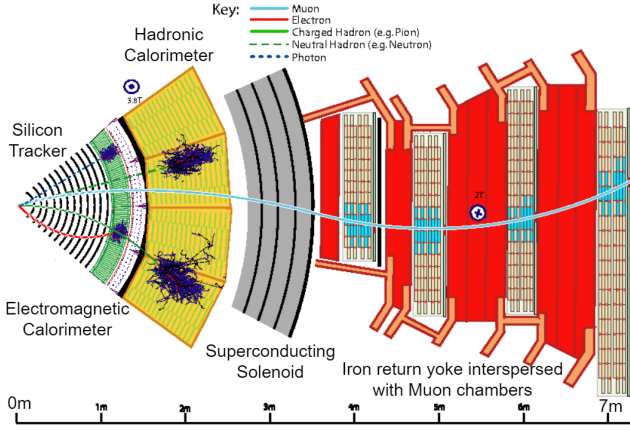


Figure 2: A transverse slice of the CMS detector, with lines representing the detection of different types of particles. Adapted from [33].

superconducting solenoid, an estimate of the particles momentum can be made [34].

Each of the layers aims to measure the properties of different types of particles. The first layer is a silicon tracker, which measures the properties of charged particles to find the decay vertices. The electromagnetic calorimeter (ECAL) measures the properties of electrons and photons. In the hadronic calorimeter (HCAL), hadronisation occurs leading to hadronic jets in the detector. The HCAL measures the properties of these jets. Beyond these layers are a series of muon chambers, specially designed to detect muons [34].

The CMS uses the information measured to attempt to reconstruct all particles in an event using the 'particle-flow' (PF) algorithm [33]. Once the particles are reconstructed and identified, the PF algorithm then calculates further properties of the events, such as missing transverse energy (MET). The data in this research was simulated based on events at the CMS detector.

### 3.2 Data Simulation

The data used in this project was produced by Monte Carlo simulations. Data generation starts by generating the underlying processes. This is done by taking the leading order or next-to-leading order Feynman diagrams to produce the desired final states, and inputting these into POWHEG or MADGRAPH5 aMC@NLO. Next, hadronisation and showering are simulated, turning any outgoing quarks and gluons into realistic jets. Geant4 is then used to simulate the particles interacting with the detector. To mimic the pile-up effect of real data, the results of additional pp collisions occurring at the same beam crossing are overlaid. The detector trigger and data collection are then simulated, before finally applying the PF algorithm to attempt to identify the underlying particles and calculate their properties [35].

Event selection criteria are then used to select whether an event is suitable for analysis, by comparing the event properties with a series of event quality cuts.

### 3.3 Event Selection

In the pp collisions at the LHC, data is produced at a rate of roughly one petabyte per second [18]. Not all of the data from these collisions can be stored, therefore trigger systems are initially put in place to select which events are discarded. At the CMS, the level-1 trigger system reduces the event rate from roughly 1 GHz to 100 kHz. The high-level trigger then further reduces this number to 1 kHz, which is the maximum rate that data can be stored and analysed at the CMS. The high-level trigger performs an initial event reconstruction [35].

The variables describing each event can be split into two categories, event-level and jet-based variables. Event-level variables are quantities calculated for an entire event, such as total transverse momentum (HT). These provide an overall characterisation of an event. Jet-based variables are quantities calculated for each individual jet in an event, such as the jets momentum. A description of the event-level and jet-based variables used in this research, along with their symbols used below, can be found in appendix A. The signal and background signatures of two event-level variables are observed in figure 3.

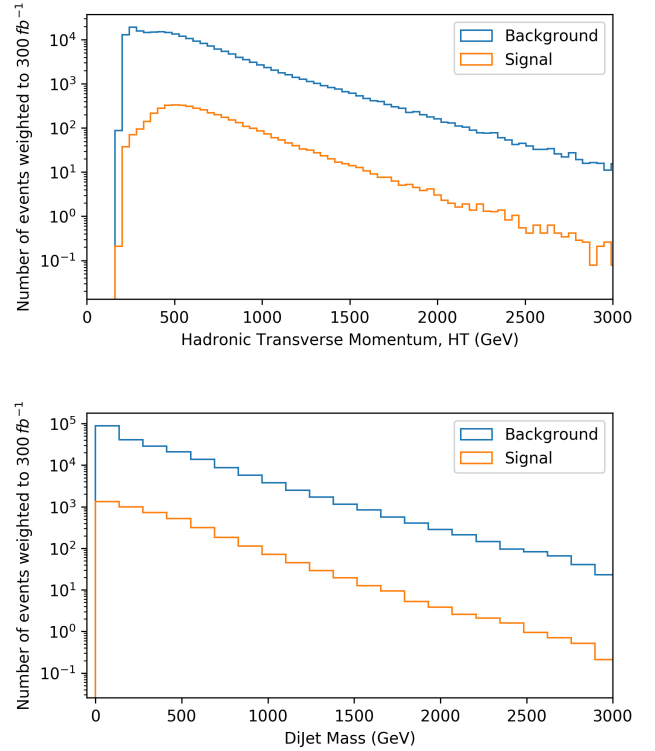


Figure 3: Distributions of the hadronic transverse momentum (upper plot) and dijet mass (lower plot) for signal and background processes. The number of events is weighted to represent run 3 at the LHC.

To ensure only high quality events are analysed, the following event quality cuts are implemented:

- $HT, MHT, MET > 200 \text{ GeV}$
- $MHT/MET < 1.2$
- The azimuthal angle between MHT and MET must be less than 0.5
- $Track MET > 80 \text{ GeV}$

- The azimuthal angle between MET and track MET must be greater than 1
- Leading and second leading jet energy  $> 80$  GeV
- Leading and second leading jet eta  $< 2.4$  (removes VBF signals)

A more complete description of the cuts made with reasons can be found in [14].

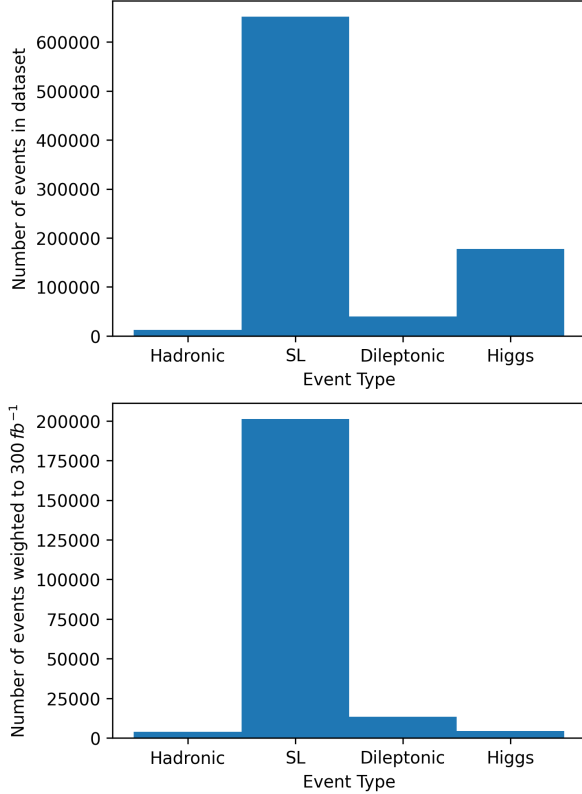


Figure 4: Distributions of the number of events in the dataset (upper plot) and the number of events weighted to  $300 \text{ fb}^{-1}$  after event quality cuts (lower plots) for each process that is evaluated in this research.

These cut significantly reduce the amount of hadronic  $t\bar{t}$  background, making semi-leptonic the dominant background, as displayed by figure 4.

The subtlety in the difference between signal and background signatures, seen in figure 3, demonstrates why machine learning is a suitable tool to identify these subtle differences and find patterns in them across multiple variables.

### 3.4 Deep Learning

Deep learning is a subset of ML that is inspired by the structure and function of the brain. Deep learning algorithms use artificial neural networks, which are made up of layers of neurons. Networks that have many hidden layers are known as deep neural networks. Every neuron in a layer is connected to every neuron in the next layer, as demonstrated in figure 5. The structure of a neural network can be generalised into 3 layers; an input layer, hidden layers and an output layer [31]. Data is passed to the input layer, and then through a series of hidden layers, before passing the result to the output layer. As the data is passed through the hidden layers, patterns are found in the data, with

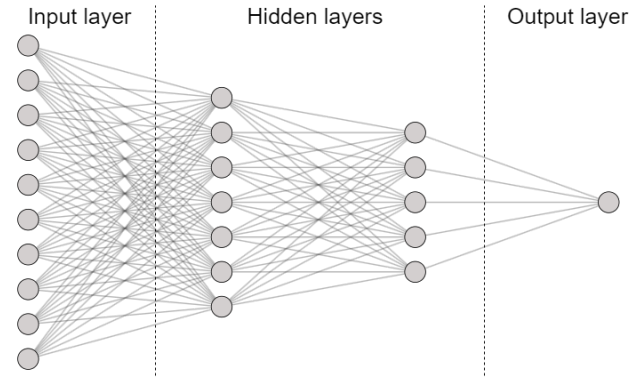


Figure 5: A visualisation of a simple neural network, displaying different layer types and connections between neurons.

each layer building on the features learned previously. In the network, the connections between neurons are weighted. The output of a neuron can be represented mathematically using the equation:

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (2)$$

where  $y$  is the output of the neuron,  $x_i$  is the inputs to the neuron,  $w_i$  is the weights of the connections between the inputs and the neuron,  $b$  represents the bias term and  $f$  represents the activation function [36].

The bias is a learned value, that shifts the output of the neuron to allow the algorithm to fit to data better and the activation function is a non-linear mathematical function that determines the neurons output.

In this research, two types of artificial neural networks are used, a feedforward neural network (FNN) and a recurrent neural network (RNN). All deep learning used the Tensorflow library in Python and was carried out on the Bristol DICE cluster, on the sc01 and GPU servers.

#### 3.4.1 Feedforward Neural Networks

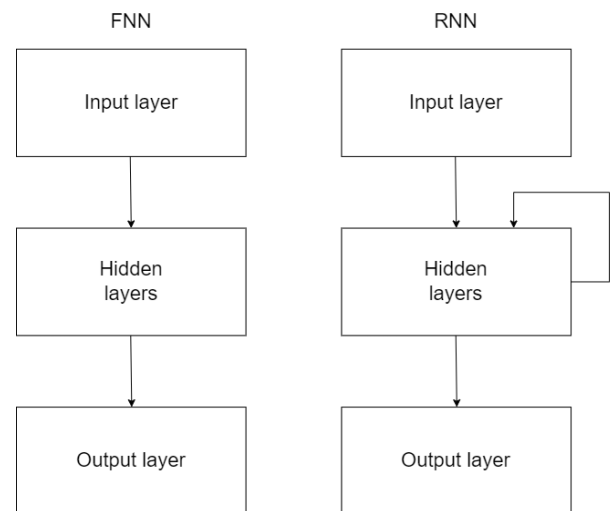


Figure 6: A simple schematic displaying the architectures of a feedforward neural network (FNN) and a recurrent neural network (RNN).



In an FNN, data only travels in the forward direction, as displayed in figure 6. Each neuron receives inputs from the previous layer, calculates a weighted sum of these inputs, adds a bias term and applies an activation function to produce an output which is passed forward to the next layer of neurons [37].

Another feature of an FNN is that all inputs must be of the same length. In this project, this means that jet-based variables can be handled in one of two ways. One option is to include each jet measurement as a new event in the dataset. The other option is to only use the information of the jet with the largest transverse momentum, known as the leading order jet. Both of these options have significant drawbacks, either creating a very large dataset or losing some information from the event entirely. This makes an FNN a suitable candidate for the analysis of event-level variables, however for the analysis of jet-based variables an RNN is more suitable. In the FNN models used in this research, the leading order jet method was implemented. Specifically, the type of FNN used in this project was a multi-layer perceptron (MLP), referring to an FNN with multiple layers.

### 3.4.2 Recurrent Neural Networks

In an RNN, the input data can be of variable length and multidimensional. This is implemented through feedback loops, which allow the output of a neuron to be fed back into itself as an input, demonstrated in figure 6. This enables information to be passed from one dimension to the next. [32].

This unique feature of an RNN makes it suitable for analysing jet-based variables, as it can use the data in its full form. The information from each jet in an event is passed through the RNN sequentially. For all RNN models used in this research, the data was padded to be the length of the maximum number of jets.

The specific type of RNN used in this research was a long short-term memory (LSTM). This type of RNN can selectively remember or forget information, which is useful in this task where padding is used, as the LSTM can forget the padded values [38].

For the most complete model, the two types of networks can be combined, enabling both the event-level and jet-based variables to have all of their information extracted. For the combination, jet-based variables are put into an RNN, turning the variable length arrays into a set of scalar values, corresponding to the predicted classification of the event. This is then funnelled into the FNN as an input variable, which captures the information from all of the jet-based variables. This model type will be referred to as the combined model, or FNN + RNN.

## 3.5 Types of Classifier

Throughout this project, a binary classifier is primarily used. In this type of classifier there are only 2 classes, signal and background. The sigmoid activation function is applied to the output layer, mapping the output values to the range of 0 to 1. Therefore, the models output represents a likelihood that an event is signal or background. In this project, 1 has been used to represent signal and 0 has been used to represent background

throughout. In the training of a binary classifier, the classification threshold, also known as the discrimination threshold or the discriminator, is set to 0.5 such that any event with a predicted value greater than 0.5 is classified as signal-like. This is only the case in training and a new optimal discrimination threshold is found once the model is trained.

A multiclass classifier is also used in this project. In this type of classifier, there can be any number of classes,  $n$ . The output layer now has  $n$  neurons and uses the 'softmax' activation function, so the output of the model is  $n$  values, representing a probability density across all classes. In training, the classification threshold is no longer a set number, instead whichever class has the highest probability is selected as the classification, but once again this is optimised once the model is trained.

In this research, a binary FNN, a binary RNN and a multiclass FNN are used.

## 3.6 Class Imbalance and Cross-section Weights

One of the challenges faced in this research is class imbalance. With the number of  $t\bar{t}H$  events being significantly smaller than its associated background processes, shown in figure 4, optimal accuracy can be achieved by only predicting background. This is a common challenge in ML, known as class imbalance. To counter this, a larger cost is applied to signal events that are misidentified, such that the model evaluates signal and background in equal proportions.

The background events are also split into their own weights. Unlike with signal and background, it is desired for the background processes to be interpreted by the model in the same ratio as they are observed at the LHC after the event quality cuts have been applied. Each event has an effective cross-section weight,  $x_s$ , representing the expected number of events in one inverse picobarn of data after the event quality cuts, such that,

$$N = \sum_{i=1}^n x_{s_i} \quad (3)$$

where  $N$  is the total number of events in  $1 \text{ pb}^{-1}$  of integrated luminosity, and  $n$  represents the total number of events in the dataset. Throughout this research, the number of events is normalised to  $300 \text{ fb}^{-1}$ , to simulate run 3 at the LHC. This is implemented by multiplying  $N$  by  $3 \times 10^5$ .

Applying the weights to the background events, the neural network evaluates the background data as 92.1% semi-leptonic, 6.2% dileptonic and 1.7% hadronic. A visual representation of this, normalised to  $300 \text{ fb}^{-1}$  is displayed in figure 4. This figure also displays the challenge of this research, showing the dominance of the semi-leptonic process in top physics at the LHC, after the event quality cuts are applied. The challenge in distinguishing event types can also be seen from the variables in figure 3, showing the difference in scale between the number of signal and background events and also displaying the similarity in the signature of signal and background.

## 3.7 Performance Metrics

In machine learning tasks, a wide range of performance metrics can be used to evaluate a model. In the training process, a

model attempts to minimise the loss function, which is a measure of how far away predicted values are from the true label. In minimising the loss function, the accuracy of the model is also generally improved [39].

Accuracy is the most intuitive metric of the success of a model, defined as

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

where TP and TN are true positive and negative respectively, and FP and FN are false positive and negative respectively [40].

While this is a key performance metric, it gives no insight into which events the model is classifying correctly and which events are being misidentified.

A confusion matrix can provide greater insights than accuracy. This is a square matrix with a row and column for each class in the classification problem, set out as follows.

True Label	Predicted Label	
	Negative	Positive
	Negative	Positive
	<i>TN</i>	<i>FP</i>
	<i>FN</i>	<i>TP</i>

Table 2: Confusion matrix configuration for a binary classification problem.

This is more insightful than the accuracy score, as it will show the amount of each class that is correctly classified and misidentified. This is particularly important when dealing with class imbalance in this research, where the TP rate needs to be optimised while minimising the FP rate. However, this metric only evaluates the model at a single discrimination threshold, so other metrics are also used.

Another commonly used performance metric in binary classification tasks is a receiver operating characteristic (ROC) curve. ROC curves are a plot of the TP rate against the FP rate for different discrimination thresholds. For a multiclass classifier one-vs-rest (OvR) ROC curves can be produced, which evaluate each class individually as the positive class. The area under the curve (AUC) is a concise metric to evaluate the performance of a classifier. Statistically, the AUC is equivalent to the probability that the classifier will output a value closer to signal-like for a randomly chosen signal event than a randomly chosen background event [41]. An AUC of 1 represents perfect classification and 0.5 indicates random classification.

The final metric used to evaluate the success of a model is significance. The significance can be calculated using the Asimov approximation for median significance formula:

$$Z = \left[ 2 \left( (s + b) \ln \left[ \frac{(s + b)(b + \sigma_b^2)}{b^2 + (s + b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[ 1 + \frac{\sigma_b^2 s}{b(b + \sigma_b^2)} \right] \right) \right]^{1/2} \quad (5)$$

where  $Z$  is the significance,  $s$  is the number of signal events,  $b$  is the number of background events, and  $\sigma_b$  is the systematic uncertainty [42].

In this research, the goal is to maximise the significance. The significance is calculated at different discriminator thresholds, and the value at which the maximum significance is achieved is defined the optimal threshold. In particle physics, a significance level of the observed signal larger than  $5\sigma$  above the expected background fluctuations is required to claim a new discovery.

### 3.8 Hyperparameters

Another important factor in optimising model configuration is hyperparameter tuning. A hyperparameter is a parameter which is defined before training takes place. They determine the architecture and outputs of a model. Hyperparameter tuning can be carried out using the GridSearchCV function in scikit. In this research, the hyperparameters were scored to optimise AUC opposed to accuracy, as AUC gives a more complete representation of the model, especially in cases with class imbalance.

Multiple optimisers were tested alongside a range of activation functions for the combined model. The best combination of optimiser and activation function was found to be Adam and ReLU. Other combinations did give a slightly higher AUC, however they ultimately gave a lower significance. This configuration of optimiser and activation function was used in all models throughout this research. Other parameters were tested for each model independently and the results can be seen in table 3.

Table 3: Hyperparameters tuning results for all models.

Hyperparameter	FNN	RNN	Combined
Hidden Layers	2	2	2
Units	256	64	256
Batch Size	256	512	256
Dropout Rate	0.2	0.1	0.2
Learning Rate	0.00005	0.0005	0.0001

For the FNN and combined (FNN+RNN) models, these hyperparameters yielded a marginally worse AUC than an optimal model, however these models were much more computationally efficient. Therefore, the marginal loss in AUC was outweighed by the increase in computational efficiency.

## 4 Results

Throughout the project, 704876 events are used in training and 176220 events are used in testing. All models are trained for 15 epochs, as this allows model saturation without overfitting. Models are also trained on 10 different random seeds to ensure that the results generalised well. The standard deviation of the 10 results is used to produce a  $1\sigma$  error.

All models make use of batch normalisation and dropout rate, which are used between every layer. Batch normalisation scales the input data to be centred around zero, with a variance of one and the dropout rate represents the probability that any neuron in any layer will be deactivated. Both of these techniques help to prevent overfitting and lead to better generalisation. The architecture of all models can be found in appendix B, figure 23.

## 4.1 Feed-forward Neural Network

The FNN is trained on 10 event-level variables, seen in table 4 in appendix A.

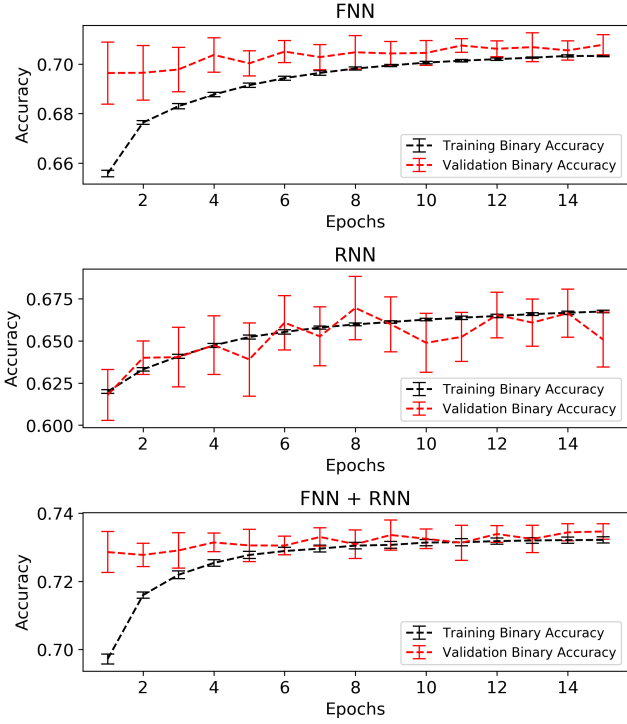


Figure 7: The average accuracy development over training for the FNN, RNN and combined (FNN+RNN) model. Error bars are  $1\sigma$ .

The FNN training accuracy curve, seen in figure 7, is characteristic of a deep learning model, showing improvement towards an asymptotic point, at which the model is saturated. Training beyond this point can lead to overfitting, where the model performs better on training data than validation data. The validation accuracy of the FNN initially varies from the training accuracy and remains slightly higher. This can occur when using dropout, as in training 20% of neurons are deactivated, but in testing all neurons are used, giving the model a higher accuracy. The optimal training accuracy is  $0.7032 \pm 0.0003$ .

ROC curves for the FNN model on different seeds were near identical, seen in figure 8. The ROC AUC is  $0.7724 \pm 0.0009$ , which shows that the model has the ability to distinguish background from signal.

A discriminator curve is a plot of the classifier's output for each class. Overlaying the test and training discriminator curves provides a good test for overfitting, as it would be expected for the test and training discriminator curves to generally be aligned, with variation suggesting overfitting. The FNN discriminator curve, seen in figure 9, is normalised so that both sets of curves are visible on the same scale. In figure 9, the double peak in background is due to the three different background processes having different characteristics. The plot also shows that the model has not overfit to training data, as both sets of curves are generally aligned. More variation can be seen in the signal curves, however this is due to the smaller sample size of signal than background, causing more random fluctuations.

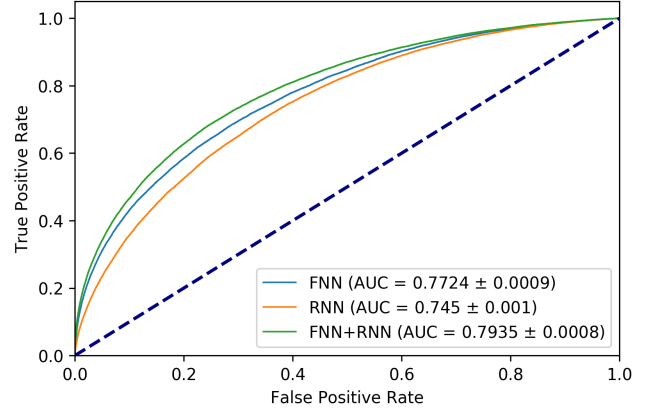


Figure 8: The receiver operator characteristic (ROC) curves for the FNN, RNN and combined (FNN+RNN) model. The blue dashed line represents the ROC curve for random classification.  $1\sigma$  error is applied to all curves, however it is not visible.

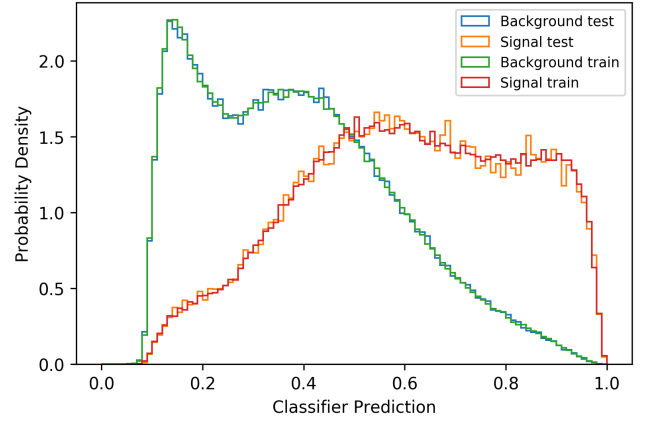


Figure 9: A normalised discriminator curve for the FNN model, where 1 represents signal-like and 0 represents background-like.

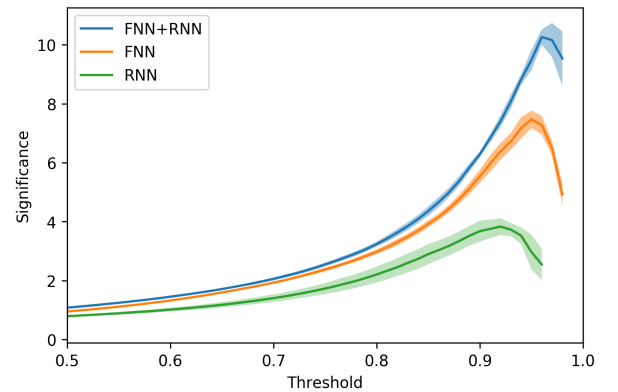


Figure 10: The significance across a range of thresholds, assuming a systematic uncertainty of 5% of the number of background events. The shaded areas represent a  $1\sigma$  error. The x-axis runs from 0.5 to 1, to focus on the key part of the plot. Before this threshold all significance curves converge to 0.

From the discriminator curves, the significance is calculated at each threshold, using equation 5, assuming a systematic un-

certainty of 5% of the numbers of background events. A significance of  $(7.6 \pm 0.4)\sigma$  is produced at a threshold of 0.95, which is the optimal threshold for this model, as displayed in figure 10.

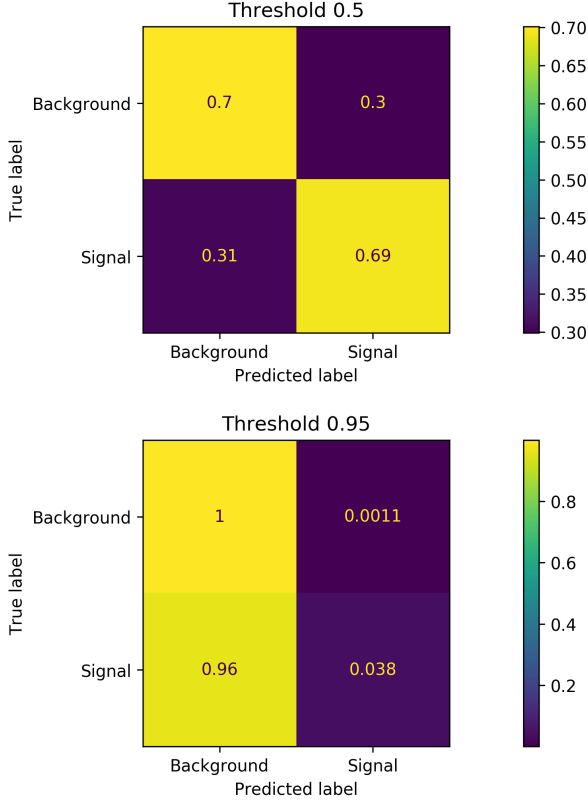


Figure 11: Confusion matrices for the FNN, with thresholds at 0.5 and the optimal threshold (0.95). The matrices are normalised to true labels (rows).

The confusion matrices in figure 11 demonstrate the significance of the threshold, showing that a higher threshold optimises the ratio of true positives to false positives. At the optimal threshold,  $160 \pm 30$  signal events and  $230 \pm 60$  background events remain, showing a significant number of dark matter events would be identified in run 3 at the LHC.

## 4.2 Recurrent Neural Network

The RNN was trained on a total of 33 jet-based variables, which can be found in table 5 in appendix A. The training curve in figure 7 shows coherence between the training and validation accuracy. This suggests that the model is not overfitting, and is generalising well to unseen data. The optimal training accuracy of the RNN is  $0.6674 \pm 0.0007$ .

As for the FNN, the RNN ROC curves from all random seeds are near identical, returning an AUC of  $0.745 \pm 0.001$ , shown in figure 8.

The normalised RNN discriminator curve, seen in figure 12, shows no signs of overfitting supporting conclusions from the accuracy curve. An initial strong spike in background predictions is observed, followed by a shallower peak. The shallow peak only drops off after 0.5, justifying the lower accuracy of the RNN, as background has many more events than signal.

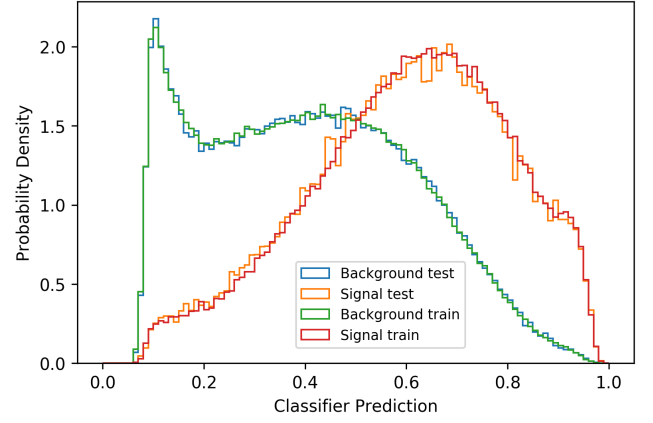


Figure 12: A normalised discriminator curve for the RNN model, where 1 represents signal-like and 0 represents background-like.

The signal curve has a much stronger peak in this model than the FNN, seen in figure 9. This suggests that using jet-based variables, the model is less confident about more events being signal than the FNN, where signal was predicted near uniformly in the range of 0.5 to 1.

The RNN significance peaks at  $(4.0 \pm 0.1)\sigma$ , seen in figure 10. This low value is expected when analysing the discriminator curve, as when signal curve peaks, the background curve is still very prominent. This optimal significance is achieved at a threshold of 0.92. At this threshold 2.7% of signal and 0.19% of background are identified as signal, corresponding to  $110 \pm 20$  signal events and  $370 \pm 60$  background events. This shows that the classification of events using the RNN model is not as pure as the event-level FNN.

## 4.3 Combined Model

The combined model (FNN+RNN) uses the same event-level variables as the FNN, but also uses RNN output as another variable, capturing all of the event-level and jet-based data into one model.

The accuracy curve over training of the combined model in figure 7 shows the initial epochs training and validation accuracy do not align, again possibly due to dropout. However, by the end of training the two accuracy curves converge, with a smaller error than the FNN and RNN. The optimal training accuracy of the combined model is  $0.7322 \pm 0.0009$ .

The ROC curves of the combined model are coherent across all random seeds, shown in figure 8. The AUC is  $0.7935 \pm 0.0008$ , suggesting that the combined model can distinguish signal from background best of the three models thus far.

The combined model normalised discriminator curve in figure 13 shows no signs of overfitting, with the general shape of both sets of curves agreeing. The initial strong background spike is present in this model. A secondary peak is also visible, however it has flattened out. The discriminator curve shows the number of classifications increasing as the true classification is approached, indicating that the model is performing well.

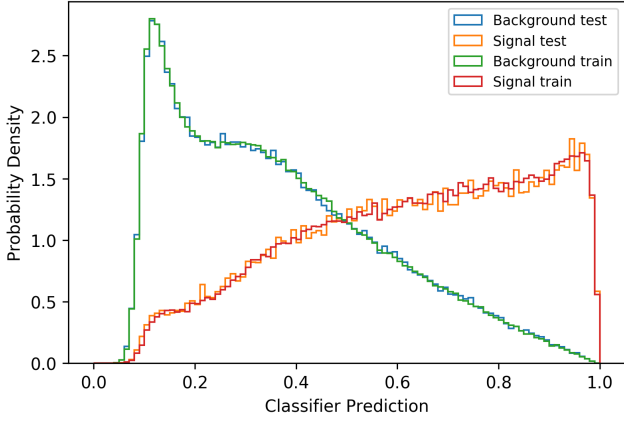


Figure 13: A normalised discriminator curve for the combined (FNN+RNN) model, where 1 represents signal-like and 0 represents background-like.

The combined model gives the highest significance, as observed in figure 10, peaking at a  $(10.4 \pm 0.4)\sigma$  at a threshold of 0.96. This agrees with the discriminator curves, as signal peaks in the 0.95 to 0.97 range, and the background is also a minimum in this range.

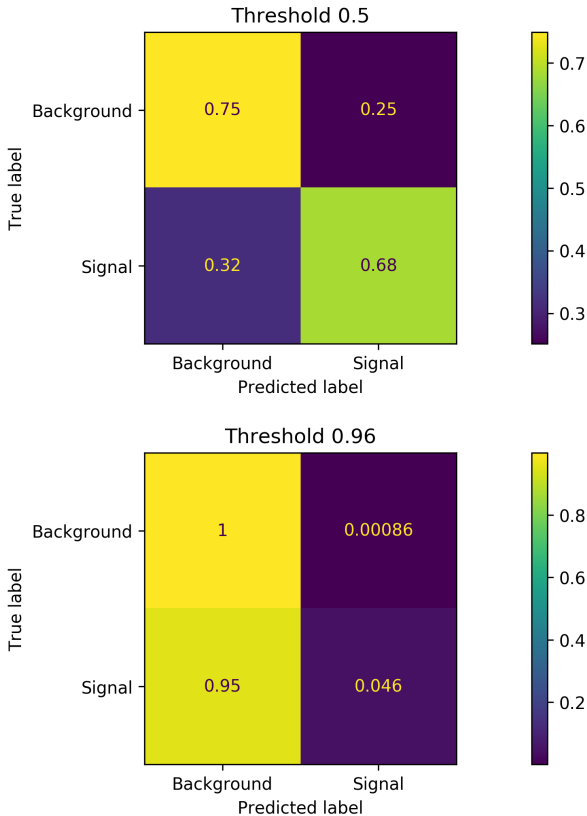


Figure 14: Confusion matrices for the combined (FNN+RNN) model, with thresholds at 0.5 and the optimal threshold (0.96). The matrices are normalised to true labels (rows).

The confusion matrix for the threshold at 0.5 and the optimal threshold can be found in figure 14. When compared with the FNN confusion matrices in figure 11, the combined model misidentifies 1% more signal, but correctly classifies 5% more background at a threshold of 0.5. This explains the combined

model having a higher accuracy score. At the optimal threshold of 0.96, the combined model identified 4.6% of signal events and 0.086% of background events as signal, corresponding to  $210 \pm 10$  signal events and  $190 \pm 10$  background events. This is a significant improvement from the FNN and RNN models in isolation, with more dark matter events being correctly classified than background being misidentified. This shows that if the Higgs to dark matter branching fraction was 100% as assumed in this research, a significant number of dark matter events would be observed in run 3 at the LHC.

#### 4.4 Multiclass Classifier

For the multiclass FNN hyperparameter tuning was not carried out, so the same hyperparameters used in the combined model were implemented. This is a reasonable assumption, as the multiclass model uses the same type of data as the combined model, just having leading order jet data and a different output layer activation function.

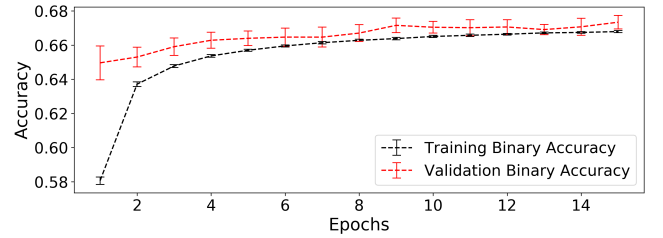


Figure 15: The average accuracy development of the multiclass classifier over training. Error bars are 1 standard deviation,  $1\sigma$ .

The training curve of the multiclass classifier, seen in figure 15, shows the initial epochs training and validation accuracy do not align, possibly due to dropout. A lower accuracy of  $0.6679 \pm 0.0006$  is obtained, which is expected as the model now has 4 classes to predict.

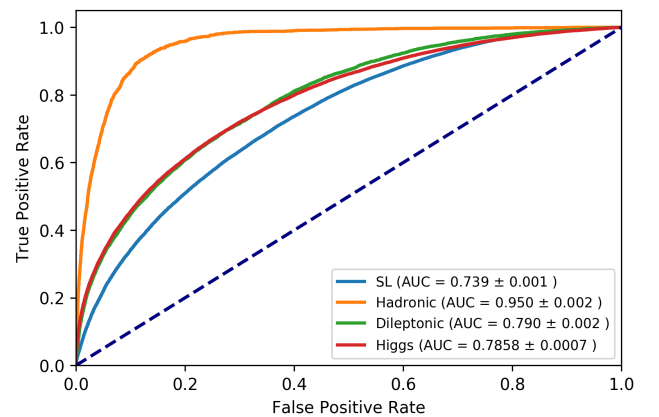


Figure 16: The one vs rest (OvR) receiver operator characteristic (ROC) curves for each class in the multiclass classifier. The positive class is labelled in the legend, where SL denotes semi-leptonic type events.

In the OvR ROC curves displayed in figure 16, the Higgs ROC curve is equivalent to the ROC curve produced by a binary classifier. The AUC results are: semi-leptonic;  $0.739 \pm 0.001$ , hadronic;  $0.950 \pm 0.002$ , dileptonic;  $0.790 \pm 0.002$ , Higgs;

$0.7858 \pm 0.0007$ . The hadronic AUC is significantly high, suggesting that the model can distinguish hadronic events very well. This is contradictory with other results in this section, so the AUC is possibly unreliable due to the very small amount of hadronic data. Due to the nature of this task, the AUC of the Higgs ROC curve is the most significant. This curve has a lower AUC than the combined model, suggesting that the multiclass model performance is not as strong.

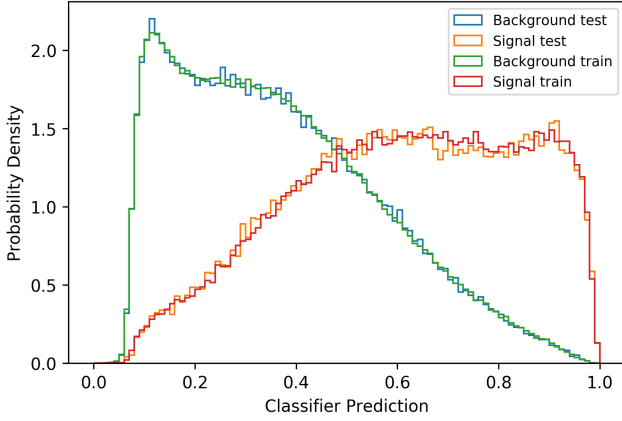


Figure 17: A normalised discriminator curve for the multiclass classifier, where 1 represents a signal-like event and 0 represents a background-like event.

For the discriminator curve, the three background classes have been grouped back into one class, making a direct comparison with the binary classifiers possible. The discriminator curve with each type of background separately, and weighted to their cross sections, can be found in figure 24, appendix C. The discriminator curve has very similar features to the FNN model in figure 9, which is reasonable due to the similarity of the two models. The discriminator curves show no signs of overfitting.

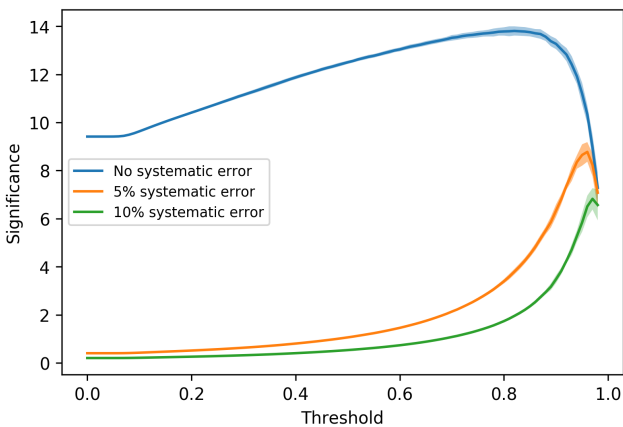


Figure 18: The significance of the multiclassifier model across a range of thresholds, with varying levels of systematic uncertainty. A  $1\sigma$  error is shaded on each line.

The multiclass model achieves a significance of  $(8.9 \pm 0.4)\sigma$  assuming a systematic uncertainty of 5% of the number of background events. This significance occurs at a threshold of 0.96, corresponding to the model being 96% confident that the events are signal-like.

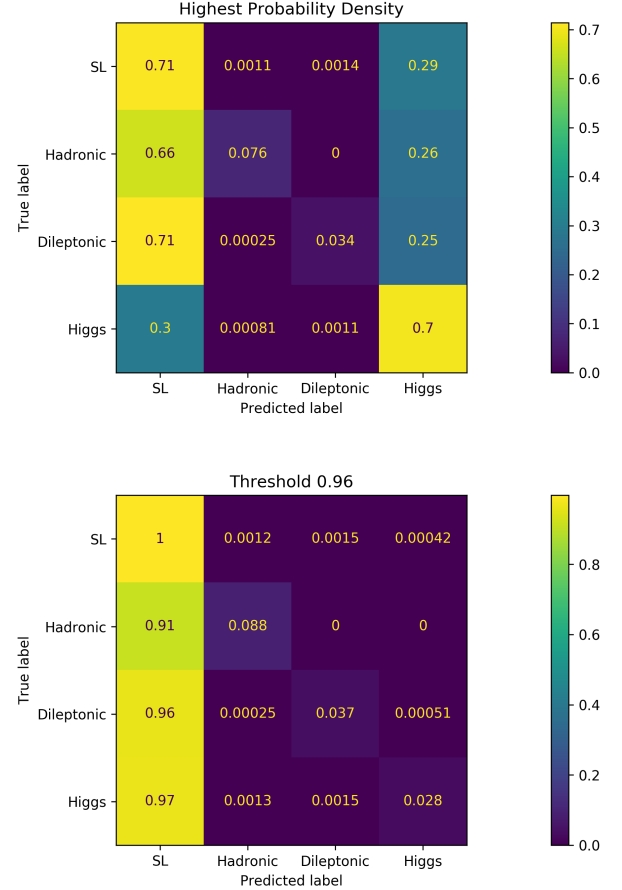


Figure 19: Confusion matrices for the multiclass classifier model, with a classification threshold of the highest assigned probability density and the optimal threshold (0.96). The matrices are normalised to true labels (rows).

From the highest probability density confusion matrix, seen in figure 19, it is observed that the multiclass model neglects the hadronic and dileptonic processes, due to their small size. They are most frequently misidentified as semi-leptonic, so overall they are mostly identified as a background process. Therefore, in the context of this task the misidentification is not an issue. At the optimal threshold, the model misidentifies a larger proportion of dileptonic events than semi-leptonic events as Higgs. However, due to the number of semi-leptonic events being an order of magnitude larger than the number of dileptonic events, semi-leptonic is still the dominant background. In total, the model identifies  $130 \pm 10$  signal and  $120 \pm 30$  background type events as signal-like. This agrees with the ROC AUC results, confirming that this multiclass classifier achieved the second best performance overall.

It is worth noting that a multiclass model would be essential going forwards, to incorporate the other dominant Higgs modes at the LHC and their associated background processes. This was the purpose of this model, however unfortunately due to time constraints this was not carried out.

## 4.5 Model Stability

This section presents an evaluation of feature importance, and model performance with varying levels of systematic uncer-



tainty.

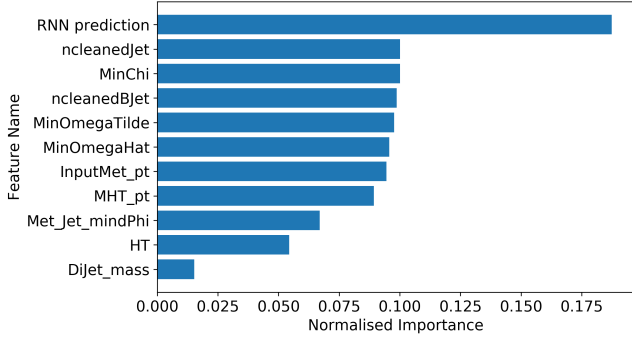


Figure 20: Distributions of the importance of each feature in the combined model, normalised such that the sum the importance of all features is equal to 1. Feature descriptions can be found in appendix A.

Figure 20 shows that the RNN prediction is the most important feature in the combined model. This is expected, as this variable encodes the data from 23 jet-based variables. Furthermore, it is the only difference between the combined model and the FNN, and the combined model shows significant improvement. Distribution of HT and di-jet mass, found in figure 3, validate the dijet mass having the least significance, as this variable has little difference in signal and background patterns. An algorithm to include whether the impact of a feature on a model is positive or negative would be a good extension here, as then these features could be dropped to improve the model.

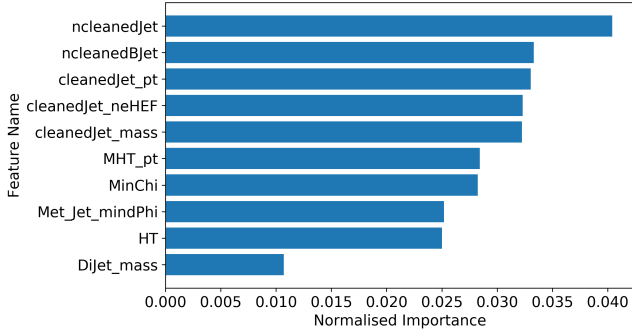


Figure 21: Distributions of the importance of the five most and least important features in the multiclassifier model, normalised such that the sum the importance of all features is equal to 1. Feature descriptions can be found in appendix A.

Only the five most and least important variables for the multiclass classifier are seen in figure 21. With no RNN prediction, the number of cleaned jets becomes the most important variable, showing consistency across models. This consistency is also seen in the low importance features, with 4 of the bottom 5 features being the same for both models. This verifies that the jet-based variables encode significant information, as these variables are more important than many event-level variables.

Figure 18 and 22 show the effect of the systematic uncertainty on significance for the multiclass and combined model respectively. It is observed that the larger the uncertainty, the significance value decreases and the optimal threshold moves towards 1. The threshold shifting towards 1 leads to fewer total events

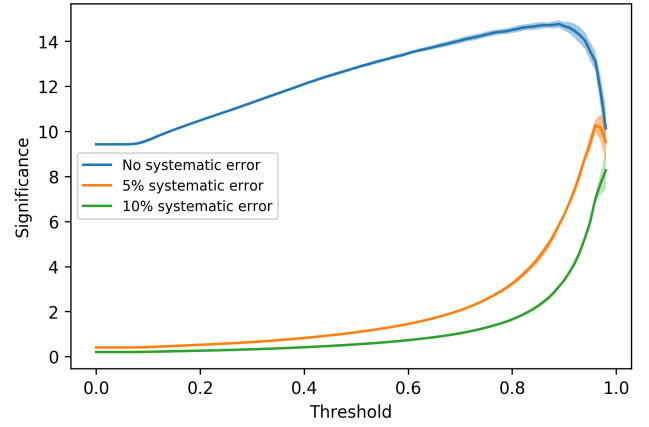


Figure 22: The significance of the combined model across a range of thresholds, with varying levels of systematic uncertainty. A  $1\sigma$  error is shaded on each line.

being identified as signal-like, therefore a higher systematic uncertainty will significantly decrease the number of dark matter events identified. This shows the importance of minimising systematic uncertainty when operating with limited datasets.

## 5 Discussion

The combined model showed significant improvement on the isolated FNN and RNN models across all performance metrics. Furthermore, the combined model showed the highest level of reproducibility, suggesting that it would be the most suitable model to use on unseen LHC data. The discrepancy in validation and training accuracy in the FNN and combined model in figure 7 is not seen in the RNN. This is consistent with the discrepancy being due to the dropout rate, as this is lower in the RNN than in the FNN and combined model. The ROC AUC of the combined model is  $0.7935 \pm 0.0008$ , displaying that the model has strong distinguishing power. This is an improvement on the FNN and RNN individually, with AUCs of  $0.7724 \pm 0.0009$  and  $0.745 \pm 0.001$  respectively.

The FNN, RNN and multiclass model discriminator curves, seen in figures 9, 12 and 17, show that these models are not highly confident in most of their classifications, with few extreme signal predictions. This is significantly improved upon in the combined model, where the peak is now close to 1, showing that the combined model has much more confidence in its signal predictions.

Another interesting feature of the discriminator curves for all models is that they do not have any events with classification values in the range of 0 to 0.05. One reason for this could be the application of class weights. The models are penalised more heavily for being further away from the true value for signal-like data, which will lead to the model being more cautious to give extreme background-like predictions, as if it is wrong the cost is amplified. At some point, 0.05 in this case, the gain from correct extreme background-like predictions is outweighed by the risk of extremely misidentifying signal.

The combined model is the optimal model due to achieving the greatest significance value of  $(10.4 \pm 0.4)\sigma$ . The error on

the combined model significance curve becomes the largest in the range 0.96 to 0.98, observed in figure 10. This is due to the number of background events being very small, while the number of signal events is large, causing the significance to have a sharp peak. The FNN produces a significance of  $(7.6 \pm 0.4)\sigma$  and the RNN gives a significance of  $(4.0 \pm 0.1)\sigma$ . The superior performance of the FNN over the RNN across all metrics is interesting, as the FNN only uses 10 event-level variables opposed to the RNN using 23 jet-based variables, which encode the data of up to 16 jets. One possible reason for this is that the RNN is not properly optimised for this task. This is a possibility, as an RNN reads the jet data sequentially, and in this RNN model the jet-based variables were fed to it in the order they were originally found in the dataset. Changing the order of the jet-based variables to have the leading order jet first and working in descending order could have helped the model learn. Furthermore, optimiser and activation function tuning was only done for the FNN models. These would be two improvements to test further in future.

The multiclass classifier’s peak significance of  $(8.9 \pm 0.4)\sigma$  shows that this is the second best model. This is consistent with expectations, as the multiclass model is given more data than the FNN, but has less encoded data than the combined model. One improvement that could be made to the multiclass model would be to create a multiclass RNN to process the jet-based data, and use its outputs as new variables in the multiclass FNN, analogous to the combined binary classifier. This would give a direct comparison between the multiclass classifier and the combined binary classifier, to ensure that the multiclass model is optimised before introducing different processes.

At the optimal threshold of each of the models, the combined model gave the best ratio of signal to background events and the most signal events, with  $210 \pm 10$  signal events and only  $190 \pm 10$  background events being identified as signal. This is 31% more signal and 17% less background than the FNN model, showing that the combined model is a significant improvement upon both independent models.

An observation of 210  $t\bar{t}H$  events with the Higgs decaying invisibly in 300  $fb^{-1}$  of data from run 3 at the LHC would provide evidence that the SM Higgs to invisible branching fraction of 0.1% is incomplete, supporting the need for BSM physics. However, the results assume a Higgs to invisible branching fraction of 100% and implementing a realistic limit, such as the upper limit set by the CMS or ATLAS, would significantly reduce the significance and the number of signal events observed.

## 5.1 Future Work

There are many different pathways for future work in this research. One extension would be to further test the stability of the models. This could be done by smearing the energies of the jets by overlaying Gaussian noise and observing how this impacts the ability of the networks to find relationships in the data. This would make the data more realistic to the experimental data collected at the LHC.

Another extension to make the model more realistic to LHC data would be to incorporate different types of signal and back-

ground processes, such as the other dominant Higgs production modes and their associated backgrounds. This would be implemented by adding them as additional classes in the multiclass classifier. The results could also be made more realistic by changing the assumption that the Higgs to dark matter branching fraction is 100%, perhaps moving it to the most stringent upper bound set by ATLAS or the CMS.

Finally, the use of different types of neural networks could be explored, such as a convolutional neural network (CNN). CNNs are trained using images, which could be constructed by flattening the detector onto a 2D plane. These images could then contain jet information, represented spatially. They have demonstrated successes in other particle physics classification tasks, so it would be interesting to see if they could extract any features from the data that were not identified by the models used in this research [43, 44].

## 6 Conclusions

In summary, a range of machine learning algorithms were used to classify top-quark events using 300  $fb^{-1}$  of proton-proton collision data at 13 TeV, simulating run 3 at the LHC. The best performing neural network was the combined model, which concatenated an FNN of event-level variables and an RNN of jet-based variables. This model achieved a peak significance of  $(10.4 \pm 0.4)\sigma$ , assuming a systematic uncertainty of 5% of the number of background events. This significance exceeds the  $5\sigma$  standard to claim a discovery of new physics. However, this also assumes a Higgs to invisible branching fraction of 100%, which exceeds the upper limit set by the CMS and ATLAS collaborations. Making this branching fraction more realistic would have a substantial impact on the significance. Under the current assumptions,  $210 \pm 10$   $t\bar{t}H$  events and  $190 \pm 10$  associated  $t\bar{t}$  background events would be identified as Higgs to dark matter decays in run 3 at the LHC.

Furthermore, a multiclass classifier achieved a significance of  $(8.9 \pm 0.4)\sigma$ . Future work would use this multiclass model to create an algorithm which can identify all Higgs production modes and give a realistic representation of physics at the LHC.



## Appendices

### Appendix A

Appendix A presents descriptions of event-level and jet-based variables used in this research. In table 4 and 5, pt denotes transverse momentum, and MET denotes missing transverse energy.

Table 4: Summary of event-level variables, with labels and descriptions.

Variable Label	Description
DiJet_mass	Invariant mass of two leading (i.e. highest pt) jets
HT	Scalar sum of all clean jet pt
Met_Jet_mindPhi	minimum delta-phi between MET and the 4 leading jets in the event
InputMet_pt	pt of MET
MHT_pt	Magnitude of vector sum all clean jet pt
MinChi	Alternative angular variable c.f. mind-phi, more detail in [45]
MinOmegaHat	Alternative angular variable c.f. mind-phi, more detail in [45]
MinOmegaTilde	Alternative angular variable c.f. mind-phi, more detail in [45]
ncleanedBJet	Number of B-tagged AK4 jets (after b-jet selection)
ncleanedJet	Number of AK4 jets (after clean jet selection)
xs_weight ( $x_s$ )	Cross-section weights of events after event quality cuts

Table 5: Summary of jet-based variables, with labels and descriptions.

Variable Label	Description
Jet_area	"Area" of jet (in eta - phi plane)
Jet_eta	Pseudorapidity of jets
Jet_mass	Mass of jets
Jet_pt	pt of jets
cleanedJet_btagDeepB	DeepB tagging algorithm's discriminator output
cleanedJet_chHEF	Charged hadron energy fraction of jet
cleanedJet_neHEF	Neutral hadron energy fraction of jet

From table 5, pseudorapidity is defined as  $\eta = -\ln(\tan\theta/2)$ , where  $\theta$  is the polar angle.

### Appendix B

Appendix B presents the architecture for all ML models used in this project.

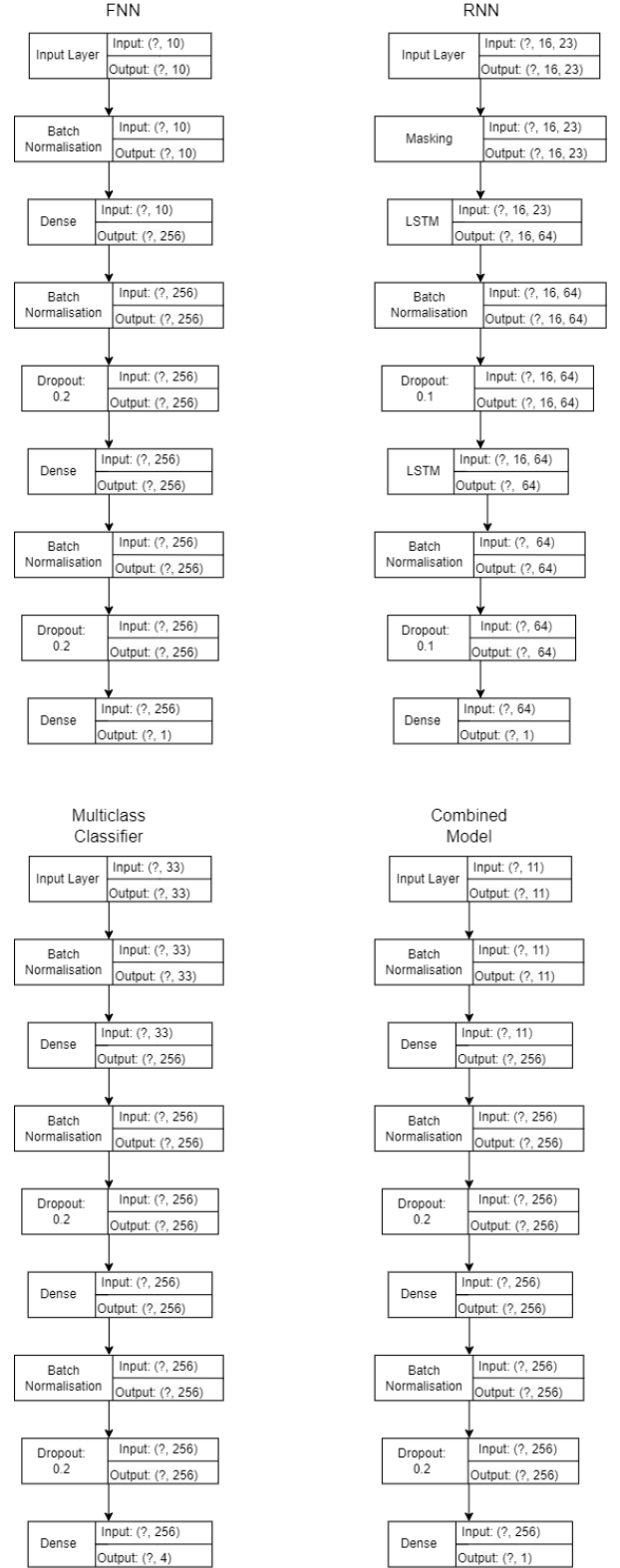


Figure 23: Simple schematics showing the architecture of all of the deep learning models used throughout this research.

## Appendix C

Appendix C presents the same multiclass classifier discriminator curve as in figure 17, but with all backgrounds separated and correctly weighted to  $300 \text{ fb}^{-1}$ .

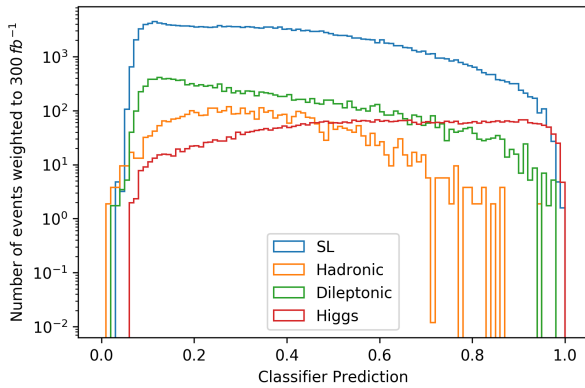


Figure 24: The discriminator curve for the multiclass model with all backgrounds separated and correctly weighted. The scale of the y-axis is logarithmic. In the legend, SL denotes semi-leptonic type events.

The figure uses a logarithmic scale on the y-axis to demonstrate the difference in scale of the number of signal and background events. Strong fluctuations towards the right of the graph for the dileptonic and hadronic discriminator curves are due to these event types having fewer simulated events in the dataset.

## References

- [1] J. D. Lykken, *Beyond the Standard Model*, CERN Yellow Report. **2**, 101-109 (2010). DOI:10.48550/arXiv.1005.1676
- [2] F. Zwicky. *On the Masses of Nebulae and of Clusters of Nebulae*, The Astrophysical Journal, **86**, 217-246 (1937). DOI:10.1086/143864
- [3] V. Rubin et al. *Rotation of the Andromeda nebula from a spectroscopic survey of emission regions*, The Astrophysical Journal, **159**, 379-403 (1970). DOI:10.1086/150317
- [4] N. Aghanim et al. (Planck Collaboration). *Planck 2018 results. VI. Cosmological parameters*, Astronomy & Astrophysics, **641**, 1 (2020). DOI:10.1051/0004-6361/201833910
- [5] D. Clowe et al. *A direct empirical proof of the existence of dark matter*, The Astrophysical Journal, **648**, L109-L113 (2006). DOI:10.1086/508162
- [6] M. Schumann, *Direct detection of WIMP dark matter: concepts and status*, Journal of Physics G: Nuclear and Particle Physics, **46**, 6-7 (2019). DOI:10.1088/1361-6471/ab2ea5
- [7] A. Arbey and M. Battaglia and F. Mahmoudi, *The Higgs boson, supersymmetry and dark matter: Relations and Perspectives*, Annalen Phys. 528 (2016) 179-186, **528**, 179 (2015). DOI:10.1088/1361-6471/ab2ea5
- [8] G. Aad et al. *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B, **716**, 1-29 (2012). DOI:10.1016/j.physletb.2012.08.020
- [9] R. Workman et al. (Particle Data Group). *Review of Particle Physics*, PTEP, **2022**, 487, 533 (2022). DOI:10.1093/ptep/ptac097
- [10] A. M. Sirunyan et al. (The CMS Collaboration). *Search for dark matter produced in association with a Higgs boson decaying to a pair of bottom quarks in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$* , Eur. Phys. J. C, **79**, 9 (2019). DOI:10.1140/epjc/s10052-019-6730-7
- [11] G. Aad et al. (The ATLAS Collaboration). *Search for invisible decays of a Higgs boson using vector-boson fusion in pp collisions at  $\sqrt{s} = 8 \text{ TeV}$  with the ATLAS detector*, Journal of High Energy Physics, **1**, 6 (2016). DOI:10.1007/jhep01(2016)172
- [12] ATLAS Collaboration, *Search for associated production of a Z boson with an invisibly decaying Higgs boson or dark matter candidates at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector*, Phys. Lett. B, **829**, 1 (2021). DOI:10.48550/arXiv.2111.08372
- [13] ATLAS Collaboration, *Combination of searches for invisible decays of the Higgs boson using  $139 \text{ fb}^{-1}$  of proton-proton collision data at  $\sqrt{s} = 13 \text{ TeV}$  collected with the ATLAS experiment*, submitted to Physics Letters B (2023). DOI:10.48550/arXiv.2301.10731

- [14] CMS Collaboration, *A search for decays of the Higgs boson to invisible particles in events with a top-antitop quark pair or a vector boson in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, submitted fevto Eur. Phys. J. C (2023). DOI:10.48550/arXiv.2303.01214
- [15] V. S. Ngairangbam et al. *Invisible Higgs search through vector boson fusion: a deep learning approach*, Eur. Phys. J. C, **80**, 2 (2020). DOI:10.1140/epjc/s10052-020-08629-w
- [16] J. Sidey-Gibbons and C. Sidey-Gibbons, *Machine learning in medicine: a practical introduction*, BMC Med Res Methodol, **19**, 64 (2019). DOI:10.1186/s12874-019-0681-4
- [17] M.D. Schwartz, *Modern Machine Learning and Particle Physics*, Harvard Data Science Review, **2** (2021). DOI:10.1162/99608f92.beeb1183
- [18] A. Radovic et al. *Machine learning at the energy and intensity frontiers of particle physics*, Nature, **560** 41-41 (2018). DOI:10.1038/s41586-018-0361-2
- [19] ATLAS Collaboration, *Measurement of the  $t\bar{t}$  production cross-section using  $e\mu$  events with  $b$ -tagged jets in  $pp$  collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS detector*, Eur. Phys. J. C, **76**, 4-5 (2016). DOI:10.1140/epjc/s10052-016-4501-2
- [20] ATLAS Collaboration, *Search for third-generation vector-like leptons in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, submitted to Journal of High Energy Physics, 4 (2023). DOI:10.48550/arXiv.2303.05441
- [21] The CMS Collaboration, *Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, Journal of High Energy Physics, **11**, 6-7 (2021). DOI:10.1007/JHEP11(2021)153
- [22] The CMS Collaboration, *Search for Higgs boson decays to invisible particles produced in association with a top-quark pair or a vector boson in proton-proton collisions at  $\sqrt{s} = 13$  TeV and combination across Higgs production modes*, CERN, CMS-PAS-HIG-21-007, 1 (2022).
- [23] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, CERN Report, **2** (2016). DOI:10.23731/CYRM-2017-002
- [24] ATLAS Collaboration, *Search for new phenomena in events with an energetic jet and missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Phys. Rev. D, **103** (2021). DOI:10.48550/arXiv.2102.10874
- [25] ATLAS Collaboration, *Search for invisible Higgs-boson decays in events with vector-boson fusion signatures using  $139\text{ fb}^{-1}$  of proton-proton data recorded by the ATLAS experiment*, Journal of High Energy Physics, **8**, 2-3 (2022). DOI:10.1007/jhep08(2022)104
- [26] CMS Collaboration, *Search for heavy resonances decaying into a vector boson and a Higgs boson in final states with charged leptons, neutrinos and  $b$  quarks at  $\sqrt{s} = 13$  TeV*, Journal of High Energy Physics, **11**, 1-6 (2018). DOI:10.1007/jhep11(2018)172
- [27] ATLAS Collaboration, *Search for a scalar partner of the top quark in the all-hadronic  $t\bar{t}$  plus missing transverse momentum final state at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Eur. Phys. J. C, **80** (2020). DOI:10.48550/arXiv.2004.14060
- [28] ATLAS Collaboration, *Search for new phenomena with top quark pairs in final states with one lepton, jets, and missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Journal of High Energy Physics, **4** (2020). DOI:10.48550/arXiv.2012.03799
- [29] ATLAS Collaboration, *Search for new phenomena in events with two opposite-charge leptons, jets and missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Journal of High Energy Physics, **4** (2021). DOI:10.48550/arXiv.2102.01444
- [30] ATLAS Collaboration, *Constraints on spin-0 dark matter mediators and invisible Higgs decays using ATLAS 13 TeV  $pp$  collision data with two top quarks and missing transverse momentum in the final state*, submitted to Eur. Phys. J. C, 16 (2022). DOI:10.48550/arXiv.2211.05426
- [31] P. Mehta, *A high-bias, low-variance introduction to Machine Learning for physicists*, Physics Reports, **810**, 4-6 (2019). DOI:10.1016/j.physrep.2019.03.001
- [32] Giuseppe Carleo et al. *Machine learning and the physical sciences*, Reviews of Modern Physics, **91**, 3-7 (2019). DOI:10.1103/revmodphys.91.045002
- [33] CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, JINST, **12** (2017). DOI:10.1088/1748-0221/12/10/P10003
- [34] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST, **3** (2008). DOI:10.1088/1748-0221/3/08/S08004
- [35] CMS Collaboration, *First measurement of the top quark pair production cross section in proton-proton collisions at  $\sqrt{s} = 13.6$  TeV*, submitted to Journal of High Energy Physics, 2 (2023). DOI:10.48550/arXiv.2303.10680
- [36] A. Krenker, *Artificial Neural Networks: Methodological Advances and Biomedical Applications*, Intech, 5 (2011). DOI:10.5772/644
- [37] D. Svozil et al. *Introduction to multi-layer feed-forward neural networks*, Chemometrics and Intelligent Laboratory Systems, **39**, 43-62 (1997). DOI:10.1016/S0169-7439(97)00061-0
- [38] M. Wielgosz et al. *Using LSTM recurrent neural networks for monitoring the LHC superconducting magnets*, Nuclear Instruments and Methods in Physics Research Section A, **867**, 43 (2017). DOI:10.1016/j.nima.2017.06.020
- [39] Q. Wang et al. *A Comprehensive Survey of Loss Functions in Machine Learning*, Annals of Data Science **9**, 187-212 (2022). DOI:10.1007/s40745-020-00253-5

- [40] A. P. Bradley, *The use of the area under the ROC curve in the evaluation of machine learning algorithms*, Pattern Recognition, **30**, 1145-1159 (1997). DOI:10.1016/S0031-3203(96)00142-2
- [41] T. Fawcett, *An introduction to ROC analysis*, Pattern Recognition Letters, **27**, 868 (2006). DOI:10.1016/j.patrec.2005.10.010.
- [42] G. Cowan et al. *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C, **1**, 26 (2011). DOI:10.1140/epjc/s10052-011-1554-0
- [43] A. R. Jimenez et al. *Convolutional Neural Networks for event classification*, PoS LHCP2021, **264** (2021). DOI:10.22323/1.397.0264
- [44] C. F. Madrazo et al. *Application of a Convolutional Neural Network for image classification for the analysis of collisions in High Energy Physics*, EPJ Web Conf, **214**, 06017 (2019). DOI:10.1051/epjconf/201921406017
- [45] T. Sakuma, H. Flaecher and D. Smith, *Alternative angular variables for suppression of QCD multijet events in new physics searches with missing transverse momentum at the LHC*, prepared for submission to JHEP (2019). DOI: 10.48550/arXiv.1803.07942

## Certification of ownership of the copyright

This Project Report is presented as part of, and in accordance with, the requirements for the degree of BSc at the University of Bristol, Faculty of Science.

I hereby assert that I own exclusive copyright in the item named below. I give permission to the University of Bristol Library to add this item to its stock and to make it available for consultation in the library, and for inter-library lending for use in another library. I also give consent for this report to be made available electronically to staff and students within the University of Bristol. It may be copied in full or in part for any bona fide library or research work. No quotation and no information derived from it may be published without the author's prior consent.

Author	Corey Swinbank
Title	Boosting dark matter searches using machine learning
Date of submission	19/04/2023

I agree that submission of this report constitutes signing of this declaration.

This project/dissertation is the property of the University of Bristol and may only be used with due regard to the rights of the author. Bibliographical references may be noted, but no part may be copied for use or quotation in any published work without the prior permission of the author. In addition, due acknowledgement for any use must be made.