Big Data: Atlas -Applications of Recurrent Neural Networks in detecting the Higgs Boson and other exotic particles

IRIS The Institute for Research in Schools

Higgs Boson: The Mass-Giving Particle

The Higgs Boson is a fundamental particle in the Standard Model and is crucial for explaining how other particles acquire mass. It gives rise to the Higgs field, which is essential for the formation of stable matter as its interactions with particles is what allows them to gain mass. It was first discovered in 2012 at the Large Hadron Collider (LHC) at CERN, a 27-kilometer ring built to accelerate protons and heavy ions to near-light speeds to collect data on their collisions. It is thanks to the silicon trackers, calorimeters and muon spectrometers in the ATLAS detector that we can detect the products of these collisions, specifically the decay products of unstable particles such as the Higgs Boson.

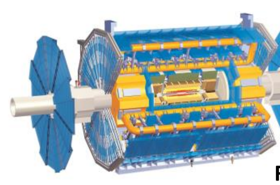
Fig 1.0 - The ATLAS detector opened [1]

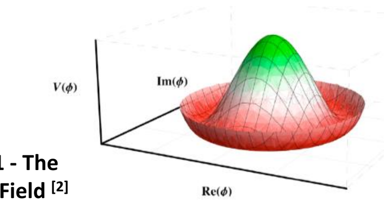
Fig 1.1 - The Higgs Field [2]

Research aims

By statistically analysing the dataset, we can detect specific particles. However, no model provides absolute certainty, and some models offer more reliable insights than others.

This brings us to our project aims:

- To **prove the existence of the Higgs boson** through 2 separate methods:
  - By **performing a non-resonant search** where we filter out background events using Monte-Carlo simulations
  - By **developing a Recurrent Neural Network** model to detect Higgs Boson events with high accuracy
- **Compare the accuracy and effectiveness** of the two models in processing the complex, sequential data generated by particle detectors
- Explore how the superior model can be used to **support the existence of theorised particles**

Through this, we hope to gain a deeper understanding of the statistical techniques that go into particle classification and how machine learning can be used to accelerate the detection of particles, as well as provide supporting evidence for hypothesised exotic particles.

From Decay Products to Detection

The Higgs Boson is almost impossible to spot directly, it simply decays too fast for us to detect. However, we can identify and measure its decay products and work backwards to piece together the particle that decayed in the first place. By looking for the right products with the right attributes, we can spot where a Higgs Boson probably decayed, thus 'finding the Higgs Boson'. Below is an example of a possible decay:

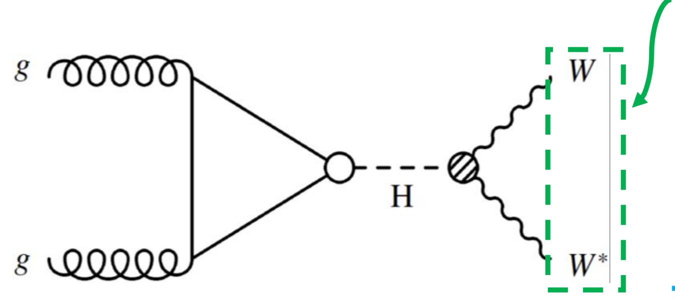

Fig 2.1 - Higgs to 2 W-Bosons channel Feynman diagram [3]

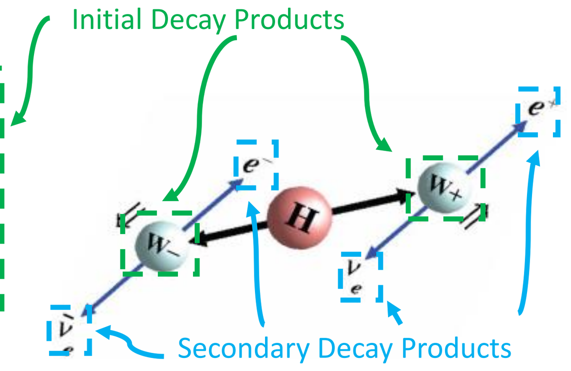

Fig 2.2 - Higgs -> W Bosons -> Lepton and Neutrino [4]

We opted to search for the Higgs Boson via the **H→ WW channel** (shown above). However, similar to the Higgs itself, W bosons decay too fast to show up in the ATLAS detector. We instead infer that they were there by detecting the Secondary Decay Products: a 'good quality lepton'(l) and a neutrino(ν). The neutrino is not detected directly, rather it too is inferred from 'missing energy'

We then find our 'good quality' lepton by imposing a set of criteria over the 600,000+ dilepton entries in order to narrow them down. This criteria is imposed by making a series of 'cuts' on our data. After this selection process, we are left with a set of secondary decay products that could have been from a Higgs Boson Decay.

However, the next layer of complexity is introduced when we consider that not every set of lν l'ν' decay products is from a Higgs Boson decay, they could have been from a quark-quark decay for example.

Conducting a non-resonant search

A non-resonant search involves looking for signals that do not produce distinct peaks or resonances in the invariant mass distribution. It is in situations like this when the decay products have **greater invariant mass** than the original particle that we must take this approach over a resonant method such as a bump-hunt.

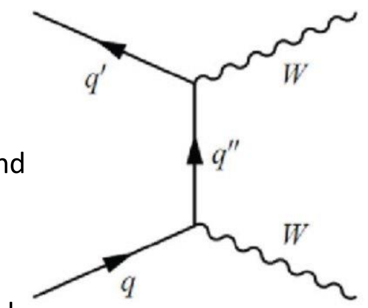

Fig 3.0 - WW-diboson background production from two quarks [5]

This approach has three main steps:

- Histogram the transverse mass of the decay products for real data
- Histogram the transverse mass of MC simulations of the background
- Subtract the backgrounds from the data

Among the most significant background events is the non-resonant **WW diboson production** stemming from 2 quarks, which we included as part of our analysis. Further background events include ττ, single-top-quark, and W⁺ jet events.

Plotting our data and then subtracting the MC data from the real data yields the following graphs. The final Histogram leaves signals remaining outside of Background events – **the Higgs signal!**

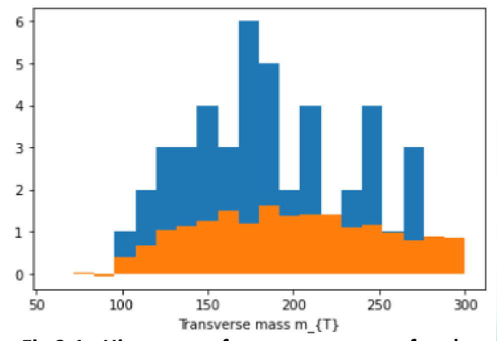

Fig 3.1 - Histogram of transverse mass of real data (blue) and MC simulations (orange)

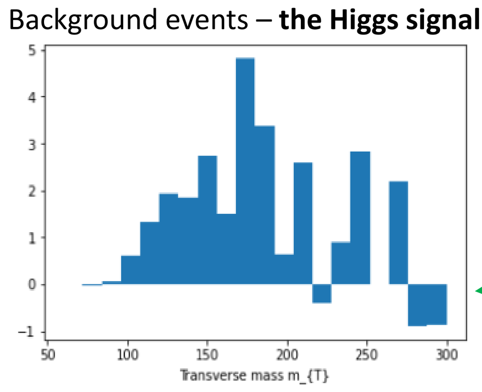

Fig 3.2 - Histogram of transverse mass of real data with simulated background events subtracted

A more reliable signal can be obtained by accounting for more background events. The following histogram accounts for less frequent events and groups them, allowing us to observe a small excess in data, and which corresponds to the **production of the SM Higgs boson.**

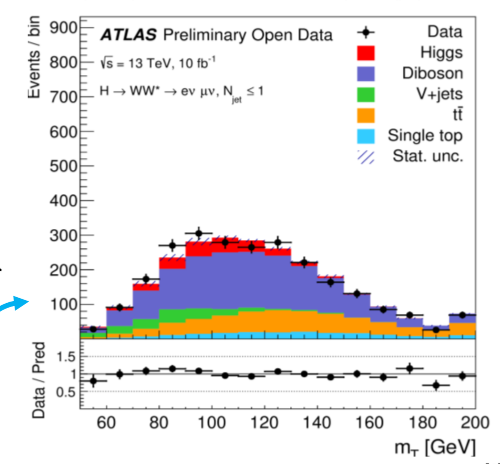

Fig 3.3 - Histogram of transverse masses, taking into account more background events [6]

What is a Recurrent Neural Network?

A Recurrent Neural Network (RNN) is a type of neural network particularly suited for processing sequential data, as it retains an internal state that captures information about previous inputs in the sequence. We can use RNNs in particle physics to analyse sequences of measurements of properties such as energy and momentum. The RNN will learn temporal patterns within these sequences, thus allowing it to recognise indicators of individual particle events

How an RNN works

An RNN processes inputs in a sequence. At each time step, ***t***, it takes an input, X_t, and combines it with the previous hidden state $\mathbf{h}_{t-1}$, using a recurrent function, ***A***.

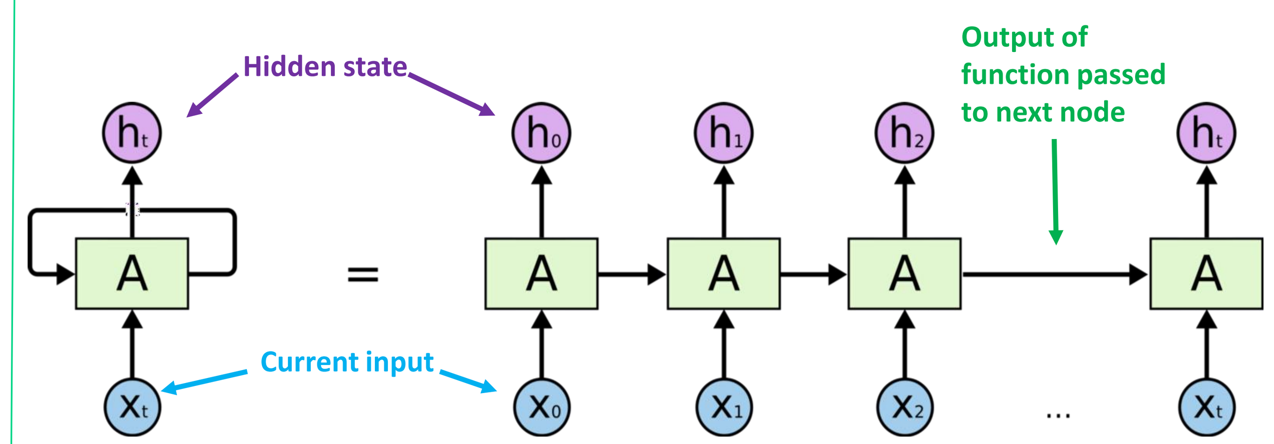

Fig 4.0 - Structure of an RNN [7]

Though RNNs are a powerful tool for sequential data processing, they suffer from the vanishing gradient problem, which limits their ability to retain information long-term. For the 600,000 events we are working with, memory can become a considerable issue and can significantly impact the result. So, we have opted to use a more advanced variation of an RNN called an LSTM.

What is an LSTM network?

A Long Short-Term Memory network is an advanced type of RNN designed specifically to overcome the vanishing gradient problem.

This occurs when the gradients used for updating the RNN's weights become very small, effectively stopping the network from further learning. In our case, would mean the RNN would struggle to capture long-term dependencies in the sequential collision data.

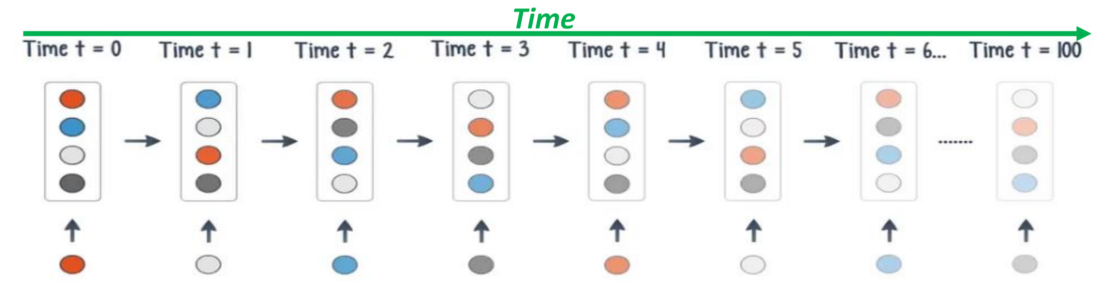

Fig 4.1 - Decay of information through time in an RNN[8]

To solve this problem, an LSTM uses LSTM units that preserve information over long periods. These units are added together in layers of an RNN and are controlled by three gates – input, output and forget – which regulate the flow of information in and out of the cell. They ensure important information is kept and irrelevant information is discarded, ensuring gradients remain stable through training.

Applying the RNN to the Higgs search

Using Recurrent Neural Networks (RNNs) to classify Higgs boson events involves a series of steps from data preparation to model training and evaluation. Our classification process involves detecting the subtle decay signatures of the particle we are looking for. For the Higgs boson, we can look at patterns such as decays into photons, W and Z bosons, or fermion pairs, amidst significant background noise.

Our actual data consists of a Training set of **250000 events**, with an ID column, **30 feature columns**, a **weight column** and a **label column** and a test set of **550000 events** with an ID column and **30 feature columns.** It contains a mixture of simulated signal and background events, built from simulated events provided by the ATLAS collaboration at CERN.

There are a couple steps in developing this:

- We first preprocess our data by **normalizing all particle features** (invariant mass, energy etc.)
- Then **employ K-Fold cross-validation**
  - This manipulates the way we use the training data to ensure robustness and avoid overfitting (when the model begins to model the training data excessively closely, instead of modelling the trends).
- **Compile the model**, including defining the necessary callbacks for the training process
- **Train the model** - this took us a *very* long time

Our model architecture consists of an **input layer** to accept sequential data, **2 LSTM layers** to process sequential information, **dense layers** to combine learned features, **dropout layers** which help avoid overfitting again, and finally the **output layer** which uses a **ReLU function** to produce a classification output.

RNN performance

We obtained the following results:
Training Accuracy – **98.7%**
Validation accuracy - **97.8%**
Test Accuracy – **97.5%**

As visible from the results, the RNN shows exceptional classification ability to separate the Higgs Boson from background noise.

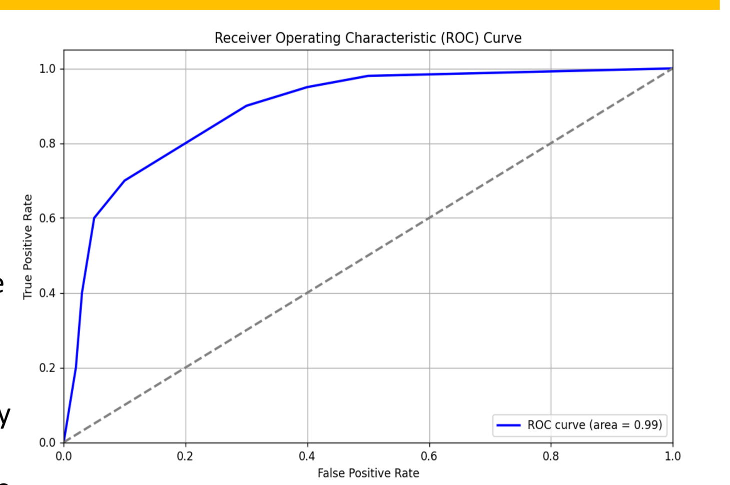

Fig 5.0 - ROC Curve of our RNN in detecting the Higgs Boson

An ROC (Receiver Operating Characteristic) curve evaluates the performance of a binary classification model by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various thresholds. It visualises the trade-off between sensitivity (how well the model detects true positives) and specificity (how well it avoids false positives). An AUC of 1 suggests perfection, whereas 0.5 suggests random guessing, so the closer the curve is to the top-left corner, the better the model's performance. The ROC curve generated from the model's predictions also showed **excellent performance** with a **higher area under the curve** (AUC=0.99).

The more effective model

There are many statistical models and methods that can be used in detecting particles, and they all have their own pros and cons depending on the situation - in our case, we are testing 2 models and looking for the model that determines the presence of the Higgs Boson most accurately and most efficiently.

Non-resonant searches typically include simpler implementation than machine learning algorithms and therefore offer a more accessible insight into the world of particle physics. However, their accuracy is very much dependant on your identification of background events and accuracy in their removal, as shown in **fig 3.2 and 3.3**. Furthermore, it is simply not possible to predict and simulate every background event, meaning there is always a considerable uncertainty attached to the output. In contrast, an RNN is much more complicated and computationally expensive to apply. However, on simulated data it has proved to be highly accurate as it does not need to consider individual background events or pre-defined cuts. Instead, it captures complex dependencies and patterns that may be indicative of Higgs boson events.

In the search of the Higgs boson, we have found using an RNN to be far superior to the tradition non-resonant search approach, especially for the **H→ WW channel** which we specifically tested.

Application in Magnetic Monopoles

Magnetic monopoles are hypothetical elementary particles that possess a single magnetic charge. The existence of a magnetic monopole in the universe would explain why electric charge is quantized. Various theories predicted them to arise naturally but haven't been observed experimentally and their discovery would have **profound implications** for our understanding of fundamental physics (unification of forces and the structure of space-time).

Magnetic monopoles are expected to leave distinct signatures in particle detectors(unique ionization patterns and long, straight tracks)- making them **perfect for Recurrent Neural Network classification**. So, we set out to evaluate the application of RNN's in the detection of elusive particles like magnetic monopoles. In our research, we utilized an RNN model to distinguish between synthetic monopole events and background noise.

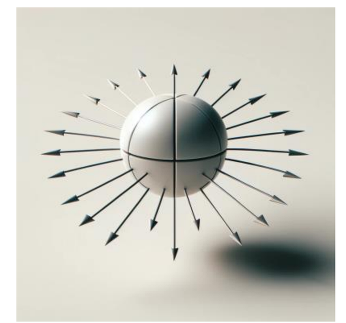
Fig 6.1 AI Interpretation of hypothesized magnetic monopoles

The exact decay channels of magnetic monopoles remain hypothetical- this makes it inaccurate to identify specific signals/signatures in experimental data like ATLAS, we instead **simulated our own data**. By generating realistic synthetic data with a Monte Carlo simulation, noise and overlapping features, we simulated monopole events (with a mass of 100 GeV, charge of 68.5, an average track length of 1000 units, and an average ionization of 100 units, while background events had a mass of 50 GeV, an average track length of 800 units, and an average ionization of 80 units) and trained our model.

The model achieved an impressive **accuracy of 90%** on test data, with precision and recall both **exceeding 90%** for both monopole and background events. The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve was approximately 0.90, indicating that the RNN effectively captured the sequential dependencies in the ionization measurements, **demonstrating high accuracy in classifying monopole events.**

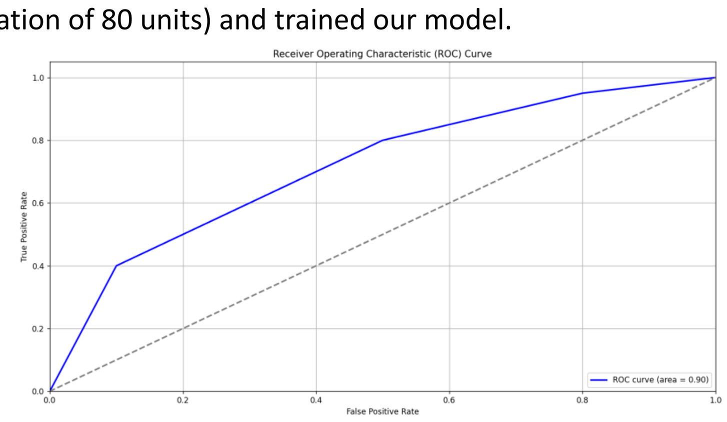

Fig 6.1 Receiver Operating Characteristic (ROC) Curve

Conclusion

The superior performance of the LSTM-based RNN model highlights the potential of advanced machine learning techniques in high-energy particle physics. Its ability to recognise patterns in temporal data suggests RNNs are particularly useful in situations where traditional methods may struggle or prove tedious due to the complexity of event sequences.

As a team we set our own personal goal at the start of this project – to build a foundation for us to apply our passion for the ever-evolving world of machine learning to the vast complexities of particle physics. Though our rigorous research and meticulous planning, time managing and organisation we feel we have achieved so much more than that and have opened our eyes to this ever-evolving field of particle discovery.

M68

Altrincham Grammar School for Boys

Harris, Nahom
Shehroz, Daniel

References:

1. ATLAS Experiment, (2008) Computer generated image of the whole ATLAS detector, CERN Document Server. Available at: https://cds.cern.ch/record/1095924 (Accessed 12/04/2024)
2. The Standard Model of Particle Physics and Beyond: https://atlasopendata.docs.cern.ch/docs/documentation/introduction/SM_and_beyond (Accessed 13/04/2024)
3. CERN accelerating science (no date) ATLAS Experiment at CERN. Available at: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2016-07/fig_01a.png (Accessed: 12/04/2024).
4. A Search for the Higgs Boson in H → WW, ,Meeting of the Division of Particles and Fields of the American Physical Society, 2011 Available at https://indico.cern.ch/event/129980/contributions/1350991/attachments/90439/129406/dpftalk.pdf (Accessed: 12/04/2024)
5. CERN accelerating science (no date) ATLAS Experiment at CERN. Available at: https://cds.cern.ch/record/1484203/files/fig1b.png (Accessed: 12/04/2024).
6. ATLAS Collaboration (2023). Example Analyses with the 13 TeV Data for Education, ATLAS Open Data Documentation. Available at: https://atlasopendata.docs.cern.ch/docs/documentation/example_analyses/analysis_examples_education_2020 (Accessed: 14/04/2024)
7. Intro to Recurrent Neural Networks LSTM | GRU, SIDDHARTH YADAV. Available at: https://www.kaggle.com/code/thebrownviking20/intro-to-recurrent-neural-networks-lstm-gru (Accessed: 15/04/2024)
8. Medium, The Vanishing Gradient Problem, Anish Singh Walia. Available at: https://medium.com/@anishsingh20/the-vanishing-gradient-problem-48ae7f501257 (Accessed: 15/04/2024)

UNIVERSITY OF OXFORD

Science & Technology Facilities Council Rutherford Appleton Laboratory