



Western Norway
University of
Applied Sciences

Search for new physics at the Large Hadron Collider

DAT255 Semester Project

Lars Erik Risholm
Marius Sellevold Hauger



Western Norway
University of
Applied Sciences



Abstract

Particle accelerators like the Large Hadron Collider (LHC) at CERN are pivotal for studying fundamental phenomena such as quark-gluon plasma and matter-antimatter differences. This project develops machine learning models to distinguish Standard Model processes from those involving the hypothetical Z' particle, aiming to detect its presence and estimate its mass. The study, conducted in three phases—training classifiers, refining models, and evaluating them on unseen data—addressed challenges like data imbalance and low recall rates. Data adjustments significantly enhanced outcomes, with neural networks demonstrating exceptional performance: classifiers achieved accuracy $>99\%$, and regression models exhibited R-squared values >0.97 . Future work should integrate ensemble methods and transfer learning to further improve these models. This research highlights the capability of machine learning to advance particle physics by enabling precise identification and mass estimation of the Z' particle.

Contents

1 Introduction	1
2 Material and methods	3
2.1 Dataset	3
2.2 Phase One - Classifier Development with Random Forest	3
2.3 Phase Two - Classifier development with Neural Networks	5
2.4 Phase Three - Parameter and Mass Estimation	5
3 Results	7
3.1 Phase One - Classifier Development with Random Forest	7
3.2 Phase Two - Classifier development with Neural Networks	8
3.3 Phase Three - Parameter and Mass Estimation	9
4 Discussion	14
5 Conclusion	17
6 References	18

1 Introduction

Particle physics investigates the foundational elements of the cosmos by examining the fundamental units of matter and their interactions. The Large Hadron Collider (LHC) at CERN (illustrated in Figure 1.1), the world's most powerful particle accelerator, together with its ATLAS experiment, is leading the investigation into these questions, including the search for the hypothetical Z prime (Z') particle.



Figure 1.1: Workers at the LHC during a long shutdown [1].

The Z' boson, which is a hypothetical particle that might exist outside the Standard Model, is thought to facilitate undiscovered physical interactions. Its detection involves identifying unique decay signatures distinct from those observed in typical standard model (SM) processes (Figure 1.2). The Z' is a gauge boson with identical couplings to the SM Z , but with a much greater pole mass, on the scale of TeV [2]. This could potentially lead to the discovery of new basic forces and provide valuable insights into dark matter and the imbalance between matter and antimatter.

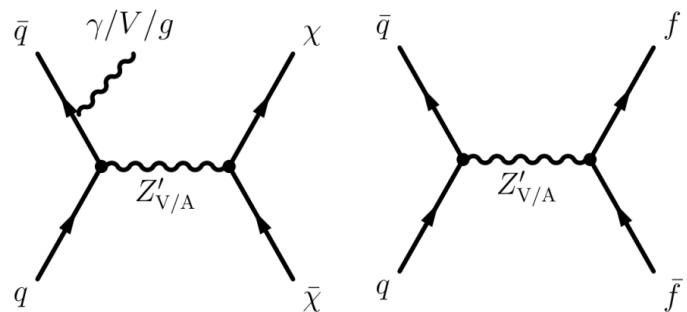


Figure 1.2: The V/A model is shown in a schematic depiction, illustrating its primary production and decay mechanisms, image from [2].

Machine learning is a subfield of artificial intelligence (AI) that enables computers to learn

from and make decisions based on data. Machine learning algorithms differ from conventional programming in that they use statistical approaches to deduce patterns and generate predictions from datasets, rather than relying on explicitly defined rules by humans. This procedure involves instructing an algorithm using a dataset, enabling it to construct a model capable of making educated forecasts or judgements without the need for explicit programming to carry out the operation.

This study utilises machine learning methodologies to identify and estimate the mass of the theoretical Z' particle by analysing data obtained from the ATLAS detector at the Large Hadron Collider (LHC). It is important to mention that while this is an individual project by Lars Erik Risholm and Marius Sellevold Hauger, it is also a part of the ongoing bachelor thesis titled "Applying Machine Learning Detection and Parameter Estimation of the Z' Particle at the LHC - Analysis Based on the ATLAS' Simulated Collision DATA (HVL)," by Hauk Hauge Mo, Lars Erik Risholm, and Marius Sellevold Hauger [3].

The primary objectives of this project are defined as:

- To develop a classifier that can accurately differentiate between signals of the Z' particle and Standard Model background processes.
- To construct a regression model capable of estimating the mass of the Z' particle from its decay products.

Finally, this project is made possible by the collaboration with the ATLAS group at Western Norway University of Applied Sciences and its associate personal.

2 Material and methods

All project details are available on Kaggle, divided into three phases. The first phase concentrated on training a random forest classifier model and exploring data features [4]. Phase two included advanced techniques for refining and validating a novel neural network classifier model [5]. In the third phase, the best model was used to test its robustness and generalizability on unseen data. Additionally, a regression model was trained and tested for mass estimation [6].

2.1 Dataset

The work employed simulated Monte Carlo datasets obtained from the CERN Open Data Portal. These datasets reflected possible interactions occurring within the ATLAS detector at the LHC. The background datasets consisted of events such as diboson and single top quark productions, while signal datasets spanned from ZPrime500 to ZPrime2500, representing Z' bosons with masses ranging from 500 to 2500 GeV/c². The databases contained information on energy, mass, and different parameters for jets and leptons. Such features and events were crucial for recognizing standard collider data, identifying, and categorizing occurrences of signal events.

These datasets were utilized for the detection of novel particle occurrences. During the third phase, the datasets for ZPrime1250 and ZPrime2250 underwent testing to verify the generalization capabilities of the model using new and previously unknown data scenarios.

Furthermore, the presence of a data imbalance, where the ratio between signal and background samples was around 1:137, required the implementation of particular strategies to address this discrepancy and enable effective model training. Data normalization and feature encoding were performed to optimize the input for machine learning models, improving performance and accuracy.

```
data.head()
```

	alljet_n	channelNumber	eventNumber	jet_1_E	jet_1_MV1	jet_1_SV0	jet_1_eta	jet_1_jvf	jet_1_m
0	0	105986	25001	54897.203125	0.059601	0.000000	0.143288	0.882837	6843.874023
1	1	105986	25003	171373.312500	0.056311	0.000000	-2.006945	0.582158	8385.192383
2	0	105986	25006	171373.312500	0.056311	0.000000	-2.006945	0.582158	8385.192383
3	2	105986	25010	735410.562500	0.053790	0.000000	-2.270990	1.000000	21263.775391
4	2	105986	25011	67439.773438	0.994451	22.765493	-0.158554	0.967006	9051.469727

Figure 2.1.1: Sample data showing features from particle collision events [7].

2.2 Phase One - Classifier Development with Random Forest

Milestone 1: Assessment of the characteristics, organization, and patterns within the dataset. Development and evaluation of three Random Forest models: a baseline model, an oversampled

model, and an undersampled model.

Data exploration

Initial assessment was conducted using Python's Pandas package to calculate descriptive statistics and generate histograms using Matplotlib for identifying data properties such as skewness and outliers. The correlations between features were effectively displayed by Seaborn heatmaps, which is important for selecting features that are not affected by multicollinearity and could potentially negatively affect the model's performance.

Analysis of jet and leptonic features was conducted to explore their relationships. Data quality was assessed by checking for missing values and inconsistencies. The process of selecting and processing features was informed by the analysis of data, in partnership with the ATLAS research group in order to identify features that distinguish between signal and background occurrences. Finally, an assessment was conducted to determine imbalances in data between the signal and background files, enabling more effective manipulation in subsequent work packages.

Classifier development

In total, three models were created using varying ratios of data (baseline, oversampling and undersampling). A random forest (Scikit-learn's 'RandomForestClassifier') approach was first selected due to its effectiveness in handling tabular data. Performance metrics for all of them were consistent, including precision, recall, F1-score and ROC-AUC.

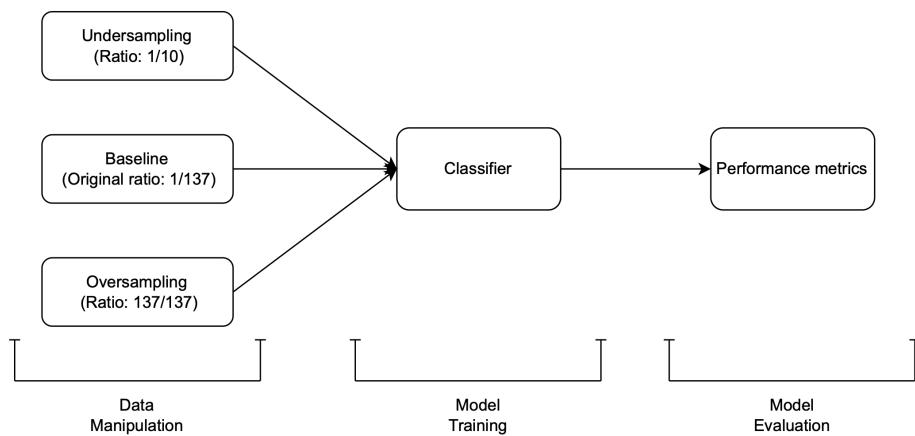


Figure 2.2.2: Schematic of the stages for training the classifier.

The Pandas package in Python was utilized to load a smaller portion (10%) of the complete dataset, which allowed for faster initial assessments and changes of the model. The data was partitioned into two distinct variables, 'df_background' and 'df_signal', based on their respective labels ('0' for background and '1' for signal samples). Ultimately, the data was rearranged using a predetermined seed value (42), guaranteeing that it can be reproduced consistently across several executions.

For both background and signal, the groups identified and removed unnecessary features, keeping a list to keep track of and maybe alter these exclusions. The data was subsequently partitioned into training and testing sets, often employing a 60/40 ratio using the 'train_test_split' function from the Scikit-learn library.

For the undersample model, a random sample of instances equals in size to the signal samples subset was selected from the background subset,, effectively balancing the two classes.

Oversampling was done using the resample function from Scikit-learn's utils package and the previously defined seed. This allowed the augmentation of the signal samples, ensuring parity with the background.

2.3 Phase Two - Classifier development with Neural Networks

Milestone 2: Development, understanding and evaluation of three different Neural networks models: a baseline model, an oversampled model, and an undersampled model..

Classifier development

On the second phase, the group switched from Random Forest to three Neural Networks models (baseline, oversampling and undersampling), taking advantage of its higher ability to capture more complex patterns and nonlinear relationships in the data.

The original dataset underwent a preliminary process, which included the categorization of categorical variables, filling in missing values, and standardizing all features. The same features that were identified in phase one were also removed from both the signal and background groups, which were thereafter divided in a 60/40 ratio. The data was consolidated into a unified data frame with a batch size of 256, in accordance with FastAI's specifications for training a neural network model.

The models were trained with an architecture of 200 layers and 100 nodes, and using the optimal learning rate obtained by FastAI's 'lr_find()' method. Similarly to phase one, performance metrics for all the models them were consistent, including precision, recall, F1-score and ROC-AUC, observed over the many epochs. Undersampling and oversampling followed the same methodology previously described.

2.4 Phase Three - Parameter and Mass Estimation

Milestone 3: Classification of new, unseen data using the single best performing classifier from phases one and two. Training of a neural network regression model capable of estimating the Z' particle mass.

Classifier evaluation

New signal datasets (500 and 2500 GeV) were loaded following the same methodology described

in phases one and two, ensuring the original signal-to-background ratio of 1/137 and evaluated using the best performing model from phase two.

Development of a regression model

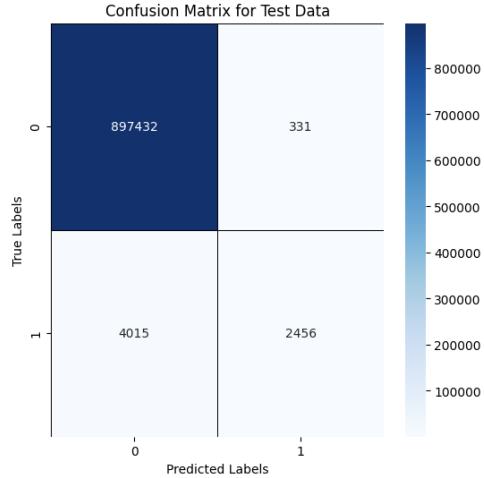
For the development of a regression neural network, the data was splitted using a different 80/20 ratio, and the pre-defined fixed seed. The network was trained using 500 layers, 200 nodes and an initial batch size of 256 for 50 epochs and a loss of $1e^{-3}$.

The model was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2).

3 Results

3.1 Phase One - Classifier Development with Random Forest

a) Baseline

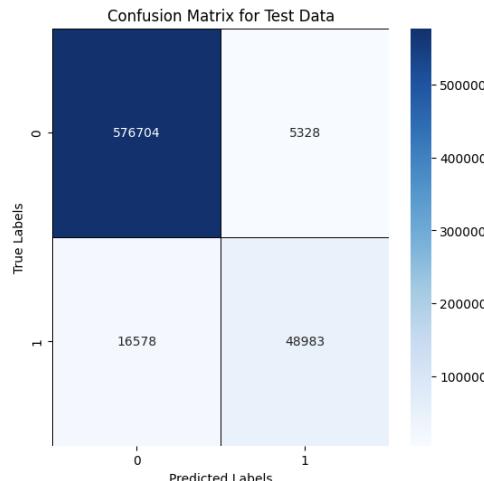


Performance metrics	
Metrics	Performance
Accuracy	0.995
Precision	0.881
Recall	0.380
F1-Score	0.531

Figure 3.1.1: Confusion matrix created for the baseline Random Forest Classifier showing number of: true positives (2456), false positives (331), false negatives (4015) and true negatives (897432).

Table 3.1.1: Performance metrics for a Random Forest Classifier trained with an imbalance of 1:137.

b) Undersampling



Performance metrics	
Metrics	Performance
Accuracy	0.966
Precision	0.902
Recall	0.747
F1-Score	0.817

Figure 3.1.2: Confusion matrix created for the undersampled Random Forest Classifier showing number of: true positives (48983), false positives (5328), false negatives (16578) and true negatives (576704).

Table 3.1.2: Performance metrics for a Random Forest Classifier trained with an imbalance of 1:10.

c) Oversampling

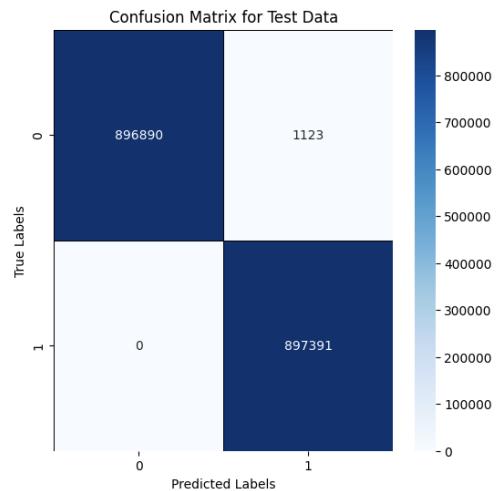


Figure 3.1.3: Confusion matrix created for the oversampled Random Forest Classifier showing number of: true positives (897391), false positives (1123), false negatives (0) and true negatives (896890).

Performance metrics	
Metrics	Performance
Accuracy	0.999
Precision	0.999
Recall	1.000
F1-Score	0.999

Table 3.1.3: Performance metrics for a Random Forest Classifier trained with an imbalance of 137:137.

3.2 Phase Two - Classifier development with Neural Networks

a) Baseline

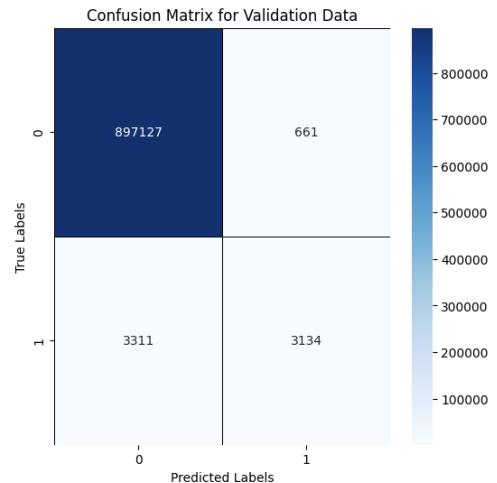


Figure 3.2.4: Confusion matrix created for the baseline Neural network classifier showing number of: true positives (3134), false positives (661), false negatives (3311) and true negatives (897127).

Performance metrics	
Metrics	Performance
Accuracy	0.996
Precision	0.826
Recall	0.486
F1-Score	0.612

Table 3.2.4: Performance metrics for a Neural network classifier trained with an imbalance of 1:137.

b) Undersampling

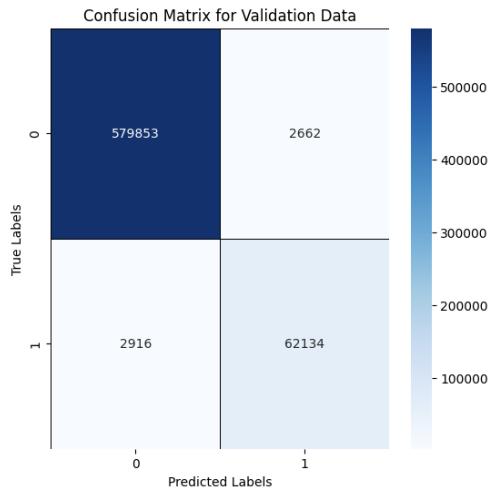


Figure 3.2.5: Confusion matrix created for the baseline Neural network classifier showing number of: true positives (62134), false positives (2662), false negatives (2916) and true negatives (579853).

Performance metrics	
Metrics	Performance
Accuracy	0.991
Precision	0.959
Recall	0.955
F1-Score	0.957

Table 3.2.5: Performance metrics for a Neural network classifier trained with an imbalance of 1:10.

c) Oversampling

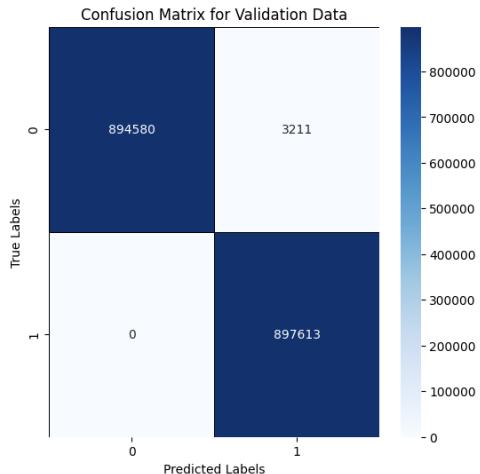


Figure 3.2.6: Confusion matrix created for the baseline Neural network classifier showing number of: true positives (897613), false positives (3211), false negatives (0) and true negatives (894580).

Performance metrics	
Metrics	Performance
Accuracy	0.998
Precision	0.996
Recall	1.000
F1-Score	0.998

Table 3.2.6: Performance metrics for a Neural network classifier trained with an imbalance of 137:137.

3.3 Phase Three - Parameter and Mass Estimation

a) Parameter Estimation through Classification

The results of the undersampled neural network classification model used for parameter

estimation across various Z' particle mass scenarios are summarized in Table 3.3.7, which presents performance metrics and significance calculations.

Z' Mass (GeV)	Accuracy	Precision	Recall	F1-Score	Significance
500	0.9986	0.9650	0.8304	0.8926	1.7751
750	0.9967	0.9550	0.5965	0.7343	1.3605
1000	0.9965	0.9551	0.5865	0.7268	1.4171
1500	0.9972	0.9573	0.6523	0.7759	1.4579
1750	0.9976	0.9582	0.6979	0.8076	1.5283
2000	0.9980	0.9582	0.7331	0.8306	1.5462
2500	0.9982	0.9560	0.7510	0.8412	1.5068

Table 3.3.7: Accuracy, precision, recall, F1-score, and significance for Z' masses ranging from 500 GeV to 2500 GeV.

Detailed confusion matrix results for each mass scenario are provided in Table 3.3.8.

Z' Mass (GeV)	TN	FP	FN	TP
500	3205385	700	3939	19285
750	3205382	703	10091	14916
1000	3205382	703	10551	14967
1500	3205382	703	8395	15746
1750	3205382	703	6997	16116
2000	3205382	703	5871	16122
2500	3205382	703	5063	15269

Table 3.3.8: True Negatives (TN), False Positives (FP), False Negatives (FN), and True Positives (TP), for Z' particle mass scenarios ranging from 500 GeV to 2500 GeV.

Performance metrics	
Metric	Value
Mean Squared Error	98.152
Root Mean Squared Error	9.907
Mean Absolute Error	4.905
R-squared	0.9998

Table 3.3.9: Updated performance metrics for a neural network regression model trained on a dataset imbalance of 1:137.

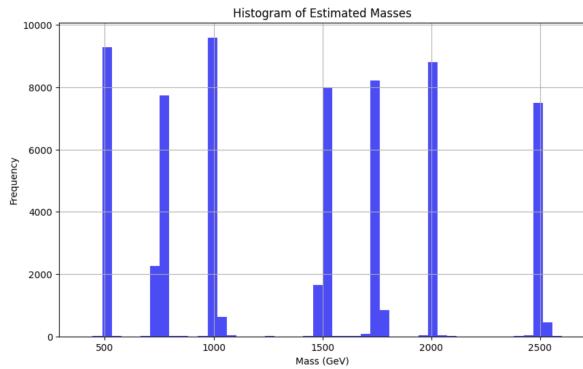


Figure 3.3.7: Frequency distribution of estimated masses by the regression model, specifically the test evaluation set containing only signal files.)

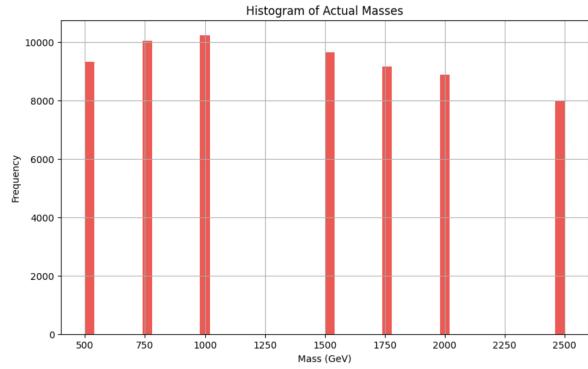


Figure 3.3.8: The actual mass parameters used in the dataset, ranging from 500 GeV to 2500 GeV, which the regression model aimed to estimate.

b) Testing the regression model on an unseen dataset

Performance metrics	
Metrics	Performance
Mean squared	52370.143
Root mean squared error	228.845
Mean absolute error	205.9
R-squared	0.789

Table 3.3.10: Evaluation of the neural network regression model on unseen data, detailing the mean squared error, root mean squared error, mean absolute error, and the R-squared value.

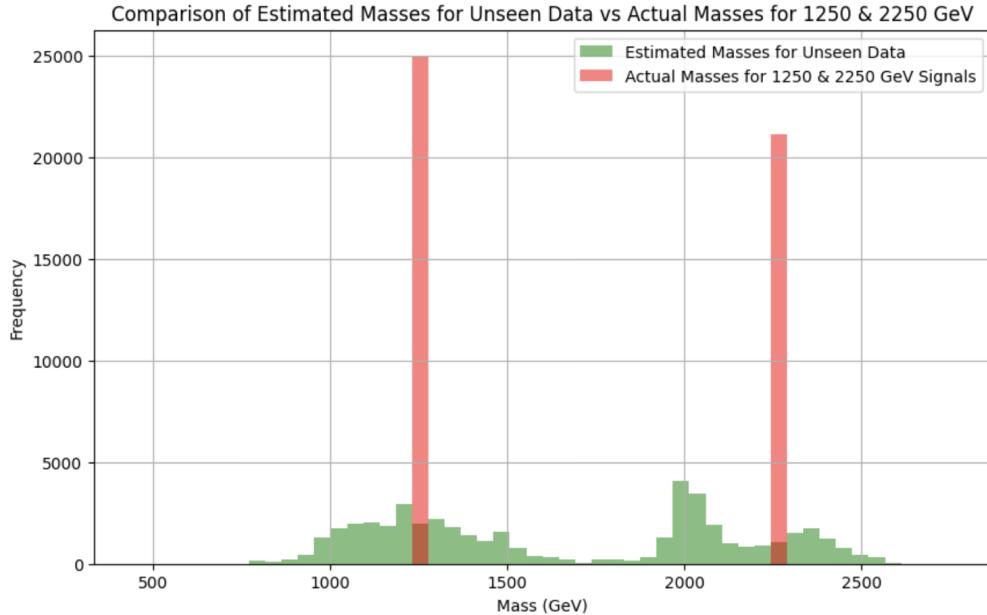
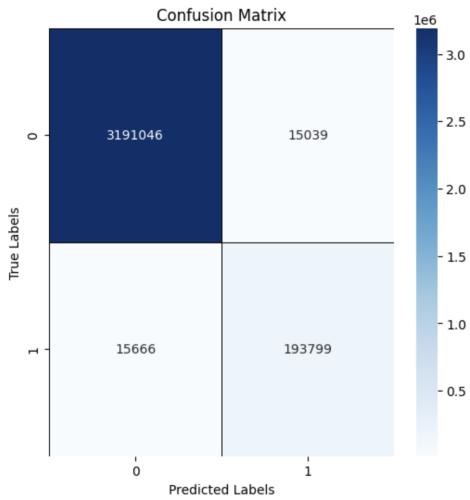


Figure 3.3.9: Comparison of the estimated masses for unseen data (in green) with the actual masses for the Z' particle signals at 1250 and 2250 GeV (in red).

c) Evaluating the best performing classifier and the newly created regression model on the whole dataset (trained + unseen).

i) Classification



Performance metrics	
Metrics	Performance
Accuracy	0.991
Precision	0.928
Recall	0.925
F1-Score	0.926
Significance	3.902
Sigma level	Strong evidence

Figure 3.3.10: Confusion matrix created for the undersampled Neural network classifier, testing on the whole dataset showing number of: true positives (193799), false positives (15039), false negatives (15666) and true negatives (3191046).

Table 3.3.11: Accuracy, precision, recall, F1-score, significance, and sigma level of the undersampled Neural Network classifier

ii) Regression

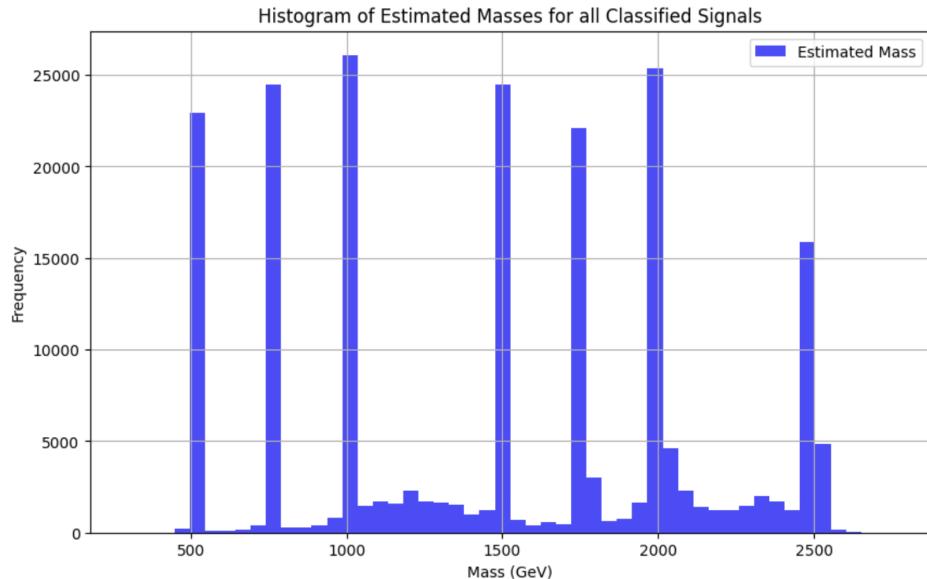


Figure 3.3.11: Frequency distribution of estimated masses by the regression model, after filtering predicted signal (true positives and false positives).

Performance metrics	
Metrics	Performance
Mean squared	10401.485
Root mean squared error	101.988
Mean absolute error	40.345
R-squared	0.975

Table 3.3.12: Evaluation of the neural network regression model on the whole dataset (trained + unseen), detailing the mean squared error, root mean squared error, mean absolute error, and the R-squared value.

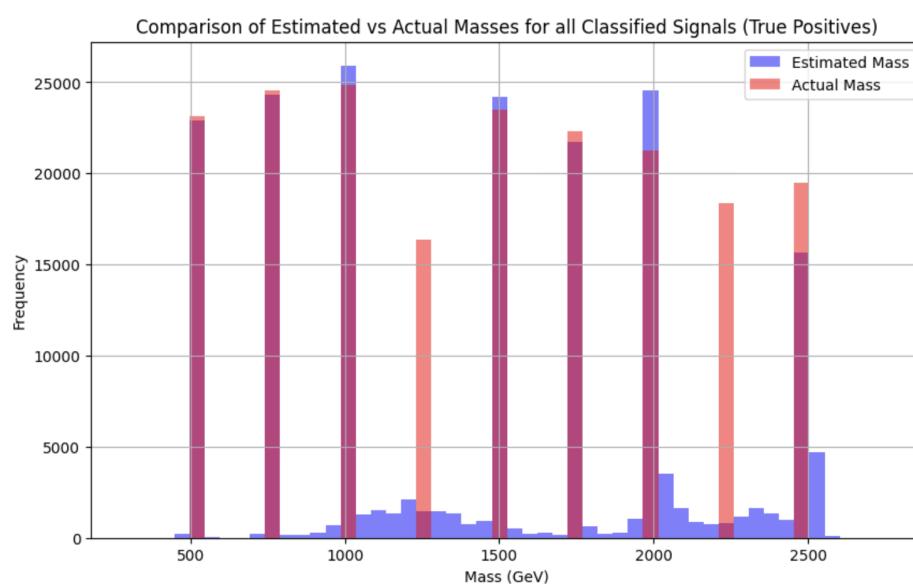


Figure 3.3.12: Comparison between estimated masses for all predicted signals (including false positives) and the actual mass parameters

4 Discussion

The objectives of this research were to develop machine learning models capable of accurately detecting the Z' particle and to employ regression models to estimate the particle's mass based on the characteristics of its decay products.

a) Classification

Comparing random forest to neural networks, both models trained on the baseline signal-to-background ratio suffered from low recall values, having good background detection but very poor signal detection, as it was expected. Oversampling the signals drastically improved the performance metrics, but were deemed to suffer from overfitting primarily due to the excessive duplication of signal data relative to the background data. Such overfitting typically results in high accuracy during training but poor generalization to new, unseen data. Undersampling the backgrounds to a 1:10 ratio improved metrics for both random forest and neural network methods, with the latter scoring slightly better with 5% better accuracy and 3% better precision. Additionally, all metrics on the undersampled neural network classifier scored higher than 95%.

b) Parameter estimation

By employing the most efficient classifier for parameter estimation tasks, we obtained useful insights into the mass-dependent behaviours of potential Z' particle signals. The classifier was tested on a variety of mass scenarios, ranging from 500 GeV to 2500 GeV. The system regularly obtained good scores on overall measures, demonstrating its robustness across various mass assumptions. Furthermore, the recently computed Sigma level indicates that although the findings are intriguing, further exploration and improvements are necessary to satisfy the 5σ benchmark for discovery in particle physics.

As the mass of the Z' particle increases, the occurrence of related events decreases, posing a greater challenge for detecting signals. Particles possessing greater masses generally exhibit diminished cross-sections, resulting in a reduction in observable occurrences. The infrequent presence of events with higher mass may impair the classifier's ability to acquire discerning characteristics that are resilient to background noise. Furthermore, when the mass rises, the decay signatures may undergo alterations, necessitating a more complex model or refined characteristics to precisely represent these subtleties.

These findings align with the fundamental principles of particle detection in the field of physics. Despite the current proficiency of the features and model architecture in capturing wider signal patterns, it may be necessary to make improvements in order to effectively detect and characterise more subtle signals that depend on mass.

c) Mass Estimation

The regression model employed to forecast the mass of the Z' particle exhibited impressive performance, accurately estimating the mass by utilising characteristics derived from particle collision events and consistently retaining accuracy in its estimations. The model successfully distinguishes between alternative mass hypotheses when applied to novel data, but suffers from a considerable decline in performance for unknown mass points (1250 GeV and 2250 GeV), as expected, with approximately 42 times larger mean absolute error and a 15% lower r-squared value.

The inclusion of three new variables related to the particle's invariant mass and its decayed state, significantly increased the models performance, especially on the unknown mass points. The chosen variables (as described in equations 1-4) were intentionally selected to improve the accuracy of predicting the mass of the Z' particle by utilising the properties of its decay products. Our goal was to compute the invariant mass in order to estimate the particle's decay state.

$$\left(\sum_{n=1}^6 \text{jet}_n E \right)^2 - \left(\sum_{n=1}^6 \frac{\text{jet}_n pt}{\text{jet}_n \cos(\theta)} \right)^2$$

Equation 1: Total squared energy minus total squared momentum for six jets, adjusting for the cosine of their emission angles, illustrates the balance of energy distribution necessary for initial state calculations.

$$\left(\left(\sum_{n=1}^4 \text{jet}_n E \right) + \text{lep}_1 E \right)^2 - \left(\left(\sum_{n=1}^4 \frac{\text{jet}_n pt}{\text{jet}_n \cos(\theta)} \right) + \frac{\text{lep}_1 pt}{\text{lep}_1 \cos(\theta)} \right)^2$$

Equation 2: Summation of squared energies and momenta for four jets plus one lepton, factoring in angle corrections, to evaluate system stability and particle interactions.

$$\left(\sum_{n=1}^2 \text{jet}_n E + \text{lep}_n E \right)^2 - \left(\sum_{n=1}^2 \frac{\text{jet}_n pt}{\text{jet}_n \cos(\theta)} + \frac{\text{lep}_n pt}{\text{lep}_n \cos(\theta)} \right)^2$$

Equation 3: Energy and momentum squared differences for two jets and leptons, providing insights into the lesser-known mass-dependent decay pathways.

$$X_n \cos(\theta) = 2 * \arctan(e^{-X_n \eta})$$

Equation 4: Relationship between the cosine of the emission angle and the pseudorapidity for jets, which is crucial for correcting particle trajectory computations.

These findings showcase the strength of machine learning in theoretical and experimental particle physics, and also open doors for future progress in the field. This study establishes a precedent for using advanced computational techniques to enhance the understanding and exploration of unknown elements within the Standard Model and beyond by effectively addressing our research questions.

Our research is in line with earlier studies that have focused on issues related to binary classification. We believe that our approach, however humble, makes a valuable contribution to the progress of this field. This study diverges, however, from prior joint collaborations that investigated image-based classification techniques or bachelor theses with comparable objectives.

It is important to acknowledge that this investigation encountered several limitations. Interpreting high-energy particle physics data was challenging without specialized knowledge in the field. Furthermore, the selection of machine learning algorithms might have been biased, which could distort the results. Moreover, the disparity in the content of the dataset, namely the abundance of Z' particle signals compared to background events, added complexity to the process of training and evaluating the model. Methods such as oversampling were employed, but more efficient methods are still required. Additionally, the scalability of our approaches was limited by the large sizes of the datasets and the high processing needs.

Assessing the benefits of ensemble methods, such as Gradient Boosting, might provide useful in enhancing the performance of the classifiers as ensemble methods have shown promise in improving model robustness and generalization capabilities in several fields. Incorporating physics-based constraints on the models, such as conservation laws and symmetry properties, may improve accuracy and reliability on both classifier and regression. Finally, more testing can be done on the benefits of using a transfer learning approach, potentially accelerating the model's training time and improving its performance, especially when dealing with limited labeled data or complex feature spaces.

5 Conclusion

This study successfully utilized advanced machine learning methods to identify and calculate the mass of the hypothetical Z' particle. Our study showcased the suitability of neural networks and regression models in analyzing intricate patterns in particle physics. Despite facing challenges such as low recall rates and algorithmic biases, the utilization of undersampling and feature enhancement techniques resulted in a significant improvement in model performance.

Neural networks demonstrated superior performance in handling imbalanced datasets across different mass scenarios. The regression models accurately estimated the mass of the Z' particle and demonstrated potential for adapting to new mass points, albeit with a decreased accuracy. The models were improved by incorporating new variables that captured specific decay product characteristics, resulting in a significant enhancement in mass estimation accuracy.

Further research should prioritize the utilization of ensemble methods to enhance robustness, the integration of domain-specific knowledge through physics-informed models, and the implementation of transfer learning to improve training efficiency. This study enhances our comprehension of the Z' particle and emphasizes the important significance of machine learning in the field of particle physics.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Quillbot AI and OpenAI's ChatGPT in order to clarify and proofread the text. After using these tools/services, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

6 References

- [1] *The Large Hadron Collider / CERN*. Accessed: 12/02/2024. URL: <https://home.web.cern.ch/science/accelerators/large-hadron-collider>.
- [2] M. Aaboud et al. “Constraints on mediator-based dark matter and scalar dark energy models using $\sqrt{\{s\}}$ = 13 TeV pp collision data collected by the ATLAS detector”. In: *Journal of High Energy Physics* 2019 2019:5 2019 (5 May 2019), pp. 1–87. ISSN: 1029-8479. DOI: 10.1007/JHEP05(2019)142. URL: [https://link.springer.com/article/10.1007/JHEP05\(2019\)142](https://link.springer.com/article/10.1007/JHEP05(2019)142).
- [3] Hauk H. Mo, Lars E. Risholm, and Marius S. Hauger. “Bachelor’s thesis: Applying Machine Learning Detection and Parameter Estimation of the Z’ Particle at the LHC - Analysis Based on the ATLAS’ Simulated Collision DATA (HVL)”. In: (2024). To be published/delivered: 13/05/2024.
- [4] *Search for new physics at the LHC: Part 1/3*. Accessed: 04/04/2024. URL: <https://www.kaggle.com/code/larserikrishholm/search-for-new-physics-at-the-lhc-part-1-3>.
- [5] *Search for new physics at the LHC: Part 2/3*. Accessed: 04/04/2024. URL: <https://www.kaggle.com/code/larserikrishholm/search-for-new-physics-at-the-lhc-part-2-3>.
- [6] *Search for new physics at the LHC: Part 3/3*. Accessed: 04/04/2024. URL: <https://www.kaggle.com/code/larserikrishholm/search-for-new-physics-at-the-lhc-part-3-3>.
- [7] *ZparticleSearch - monte carlo simulation datasets*. Accessed: 15/02/2024. URL: <https://www.kaggle.com/datasets/larserikrisholm/zparticlesearch>.