

# Higgs Boson Detection

Martín Alejandro Castro Álvarez

Universidad Internacional de Valencia

Calle Pintor Sorolla, 21 46002, Valencia (España)

[martinecastro.10.5@gmail.com](mailto:martinecastro.10.5@gmail.com)

Source: <https://github.com/MartinCastroAlvarez/higgs-boson-machine-learning>

**Abstract.** The Higgs boson is the particle responsible for giving mass to matter. However, detecting it is challenging due to its rare occurrence and rapid disintegration. This paper proposes a Deep Learning approach to detect its presence in real-time from signals generated by the Large Hadron Collider at CERN. A cautious process of data cleaning, anomaly detection, dimensionality reduction, normalization, and visualization revealed a complex multi-dimensional relationship between particle collision data and the presence of this exotic particle. As a consequence, a multi-layer perceptron was implemented because of its ability to learn non-linear patterns. The resulting model guarantees an accuracy of approximately 77%, and adjusting the prediction threshold results in a precision of approximately 97%. This research not only leads to a better understanding of fundamental laws of the universe but also promises new discoveries and technological advancements.

**Keywords:** Higgs Boson, Particle Physics, Machine Learning, Large Hadron Collider, CERN, Neural Networks, Deep Learning, Data Preprocessing, Anomaly Detection, Dimensionality Reduction, Data Augmentation, Exploratory Data Analysis, Robust Estimators, Visualization, Multi-Layer Perceptron, Binary Classification, Dropout Layer, Overfitting, Precision, Recall, Accuracy, Cross-Validation, Xavier Initialization, Adam Optimizer, Early Stopping.

## 1. Introduction

### 1.1. Problem

Particles, that form us and everything around us, initially had no mass and moved at light speed, making the formation of the universe as we know it today impossible. The Higgs boson, messenger of the Higgs' field, is the secret to understanding how particles gained mass, allowing the universe to develop into its current state [1].

### 1.2. Feasibility

Machine Learning (ML) can significantly contribute to the study of the Higgs boson due to its ability to manage, analyze, and interpret the vast and complex data produced by particle collisions. ML algorithms can improve signal detection, increasing the efficiency of identifying and studying the Higgs boson's properties as well as other particles.

### **1.3. Relevance**

First, the volume of data generated in particle physics experiments, like those conducted at CERN [1], is enormous and complex. ML can assist in analyzing this data more efficiently than traditional methods. Second, the rare occurrence and rapid decay of the Higgs boson makes it a perfect candidate for the anomaly detection capabilities of modern ML techniques.

### **1.4. Impact**

Applying ML to the study of the Higgs boson can lead to revolutionary discoveries about matter, energy, and the early conditions of the universe. On top of that, there are unimaginable technological advancements that could be derived from this research, in the same way that past particle physics discovery contributed to information technology, healthcare, aerospatial navigation, and so on [1].

## **2. Data**

### **2.1. Data Source**

The dataset, downloaded from UC Irvine Machine Learning Repository [2], and referenced by Kaggle [3], presents the problem to classify 11 million particle collision events into two categories: those signals that result in the production of the Higgs Boson particle and those that do not. Each event is derived from simulated particle collisions, as observed by the Large Hadron Collider at CERN, and is represented by 28 features reflecting the trajectories of decay particles. These features include both low-level kinematic properties directly measured by the particle detectors and high-level features mathematically derived by physicists.

### **2.2. Data Cleaning**

According to common Machine Learning practices [4], the dataset cleaning step involved verifying that all features were in a numerical format, therefore eliminating the necessity for type conversion. Then, duplicate rows were removed to ensure the uniqueness, and rows with null values were eliminated. Finally, feature scaling was then applied (See Fig. 1), normalizing the data to ensure that each feature contributed uniformly to the signal of the Higgs boson.

### **2.3. Anomaly Detection**

Outliers, defined as data points extending beyond 8 standard deviations from the mean (See Fig. 2), could potentially lead to the exclusion of a significant portion of the dataset. As a result, the 99th percentile was employed as a threshold for outliers, therefore retaining more data for model training. This approach guarantees the exclusion of only the most extreme measurements, probably caused by errors or noise, from the training process.

## 2.4. Dimensionality Reduction

By presuming all features as good predictors and iteratively discarding the one that affects the R-squared variance [5] of the multi-linear regression model the least, only the best combination of predictors needs to be used during the training step (See Fig. 3). As a result, we minimize the combinatorial explosion problem [6].

## 2.5. Data Augmentation

Generative models have been unnecessary for this dataset as it has a balanced distribution between the positive and negative target category, as demonstrated by the exploratory phase. The counts of the 'signal' feature, with approximately 5.8 million records for class 1.0 and 4.8 million for class 0.0, do not let us reject the null hypothesis that the mean of the normal distribution is different than 0.5, with a 10% margin of error. Therefore, the dataset is balanced. The fact that the proportion is not exactly 50% could be entirely attributed to the sampling process.

# 3. Exploration

## 3.1. Domain Intuition

The Feynman diagram (See Fig. 4) describes the process by which two gluons fuse into a heavy but electrically neutral Higgs boson, which eventually decays to a W boson and quarks [7]. Unfortunately, a similar process might take place but without generating a Higgs boson. These processes both generate signals that are captured by detectors, and later converted to digital data.

The columns in the dataset correspond to the following physical phenomena, found in hadron colliders, such as the LHC:

1. '**signal**' which is 1 when the signal is caused by the presence of the Higgs boson, and 0 otherwise.
2. '**lepton pT**' measures the momentum of leptons (such as electrons and muons).
3. '**lepton eta**' and '**lepton phi**' are the angles of the direction, relative to the beam, in which the lepton is moving.
4. '**missing energy magnitude**' is the amount of energy that is missing, according to the law of conservation of energy, after the collision, which might have been taken away by particles that have not been detected.
5. '**missing energy phi**' indicates the direction in which the undetected particle might have escaped.
6. '**jet 1 pt**', '**jet 2 pt**', '**jet 3 pt**', and '**jet 4 pt**' indicate the momentum of jets (streams of particles coming from quarks or gluons).
7. '**jet 1 eta**', '**jet 1 phi**', '**jet 2 eta**', '**jet 2 phi**', '**jet 3 eta**', '**jet 3 phi**', '**jet 4 eta**', and '**jet 4 phi**' indicate the 2 angles that define the direction, relative to the beam, of the jets.
8. '**jet 1 b-tag**', '**jet 2 b-tag**', '**jet 3 b-tag**', '**jet 4 b-tag**' indicate whether the corresponding jet is likely to have been generated from a heavy b-quark.
9. '**m\_jj**', '**m\_jjj**', '**m\_lv**', '**m\_jlv**', '**m\_bb**', '**m\_wbb**', '**m\_wwbb**' are derived data points calculated by physicists which might help in the discovery of the presence of the Higgs boson.

### 3.2. Statistical Summary

Numerical insights were extracted from the dataset, focusing on robust statistical measures [8] that are less sensitive to outliers. For each feature, the following robust location statistics (See Fig. 5) were calculated: minimum, maximum, deciles, median, trimmed mean, and winsorized mean. In addition, the following robust variance statistics were calculated: variance, standard deviation, skewness, kurtosis, interquartile range (See Fig. 6).

### 3.3. Visualization

Stacked histograms were generated for each feature, differentiating between the presence and absence of the Higgs Boson particle (See Fig. 7). Given the binary nature of the target variable, histograms offer a more appropriate representation than scatter plots or heatmaps for visualizing the data. These charts indicate that the data is symmetrically distributed, so the dataset is suitable for the learning step.

Also, skewed features (as also indicated by skewness and kurtosis) have been re-scaled and centered, as a result of this task (See Fig. 8). Unfortunately, the centered dataset was later discarded as part of the training process due to poor accuracy, which remained well below 50% until the decision was made.

Finally, stacked linear plots were generated, indicating the accumulation of data points per decile (See Fig. 9). This confirmed that the data is very well-distributed and no further data cleaning is considered necessary.

Remarkably, the histograms and the line charts of the decile distribution guided the entire process of data cleaning. In particular, they have been of particular interest during the normalization step.

### 3.4. Hypothesis

As a result of the observations in sections 3.1, 3.2, and 3.3, the data suggests that:

1. The '**lepton pt**' feature is skewed, indicating that particle momentum tends to be low. According to 3.1, this is expected, since high-momentum particles are more rare than low-momentum particles.
2. The '**lepton phi**' distribution appears to be uniformly distributed, as demonstrated by the symmetry of the histogram. This indicates that the direction of the lepton is completely random.
3. All the '**jet b-tag**' features show a nearly boolean behavior which aligns with 3.1 with pretty much confidence.
4. The '**missing energy magnitude**' and '**phi**' features do not show a clear separation between the two classes.
5. All the '**pt**' features for jets show long-tailed distributions, confirming that results also have a higher probability of holding low energy momentum. This information could hold the key to discovering the presence of the Higgs boson.
6. The '**eta**' features indicate that most events are centrally located, indicating that the resulting angles have a normal distribution.
7. The '**phi**' features are uniformly distributed, indicating that they are completely random.

8. The high energy '**m**' features all show skewed distributions. This indicates the high-energy values have a very low probability of occurrence.
9. In the '**jet b-tag**' features, there are consistently more instances where the Higgs Boson is present than not.
10. Creating 3D (or higher-dimensional) charts for multi-dimensional analysis is impractical due to the exponential increase in combinations.

## 4. Optimization

### 4.1. Model Selection

The previous analysis has revealed multidimensional relations within the dataset that reduces the effectiveness of linear models such as linear regression. Also, while decision trees and random forests offer improved adaptability to non-linear relationships, they require biased branched factors, like using Shannon's Entropy in node splitting. As a result, a more flexible approach is required. Neural Networks, also known as "universal approximators" [9] are preferred.

### 4.2. Supervised Learning

The dataset is already labeled. Records corresponding to signals generated by the presence of the Higgs boson are well differentiated from those that correspond to its absence. As a result, a supervised learning approach is preferred.

### 4.3. Neural Network Architecture

For the Neural Network architecture, a multi-layer perceptron (MLP) is proposed (See Fig. 10), considering its efficacy in capturing complex patterns [10]. Each layer corresponds to a set of perceptrons described by (1) in which  $x$  is the input array,  $w$  is the weight matrix, and  $b$  is the bias term.

An MLP with multiple hidden layers and ReLU activation (2) functions could provide the capacity to deal with such complexity while avoiding overfitting. However, the activation function of the output layer is a sigmoid function (3), so that predictions take values between 0 and 1.

$$z = wx + b \quad (1)$$

$$a = \max(0, z) \quad (2)$$

$$a = 1 / [1 + \exp(-z)] \quad (3)$$

Hidden layers let the neural network learn patterns that remain otherwise hidden and achieve generalization.

In addition, dropout layers prevent the model from giving too much importance to the same features. Since inputs are randomly dropped (although with a very low probability) so that different patterns resulting in the detection of the Higgs boson are learned.

The binary cross entropy loss function (4) is recommended [10] for binary classification tasks, since it rewards true positives ( $y = 1, a = 1$ ), and true negatives ( $1 - y = 1$  and  $1 - a = 1$ ), while punishing false positives ( $y = 0$  and  $a = 1$ ) as well as false negatives ( $y = 1$  and  $a = 0$ ). However, since this loss is calculated on a per row basis, it needs to be averaged by (5).

$$L(y, a) = y * \log(a) + (1 - y) * \log(1 - a) \quad (4)$$

$$J = -\frac{1}{N} * \sum L(y, a) \quad (5)$$

Convolutional layers and filters are not required when dealing with a dataset of only 25 features [10]. They are more suitable for high-dimensional data, like images.

The Adam optimizer was chosen due to its automatic computation of adaptive learning rates, to avoid having to adjust the learning rate, although RMSProp and Adagrad have also been tested.

On top of that, Xavier (Glorot) initialization randomizes the initial weight matrix of each layer in order to address the problem of slow convergence.

#### 4.4. Data Splitting

The dataset has been divided into training, validation, and test datasets, according to [9]. The training set is used to train the model, the validation set is used to do hyperparameters tuning, and the test set is used to evaluate the final model so that overfitting is avoided. On top of that, a stratified sampling strategy keeps an uniform distribution of the signal and background events, so that the model is not trained or evaluated on a biased dataset.

### 5. Results

#### 5.1. Hyper-Parameter Tuning

The model's architecture was mutated many times based on the metrics generated by previous runs (See Fig. 11). To begin with, the model had fewer than 100 neurons per hidden layer, but that seems to have been inadequate to capture the complexity of the data (as indicated by a poor accuracy). Consequently, that number was increased to more than 300 neurons per layer, thus driving accuracy from an initial 40% to above 60%.

In addition, the model was originally designed with three hidden layers, as referenced in the original paper [7]. However, in the search for higher accuracy additional layers were integrated, increasing the accuracy to 75%

On different runs, overfitting was a recurring problem, as suggested by a large difference between training and validation accuracy. To address this, dropout layers were introduced, forcing the model to learn patterns rather than memorizing individual events.

Furthermore, a balance was found, due to trial and error, between the number of hidden layers and the size of each layer, leading to a significant reduction in epoch duration from 15 minutes to 5 per epoch without compromising the model's accuracy.

The same trial-and-error approach optimizing batch sizes and epochs revealed that while larger batches (exceeding 5.000) reduced training time, they could damage accuracy up to 10%. Although Grid Search was tried, it was disregarded due to computational requirements, and its poor benefits in comparison to the trial-and-error strategy.

## 5.2. Evaluation Metrics

At the end of each epoch, the validation set was used to estimate 3 metrics: Accuracy, precision, and recall. In particular, precision and recall, with the only purpose so as to evaluate whether the performance was improving on each epoch. Yet, in this particular case, there is really no difference between false positives and false negatives, as it happens, for example, in medicine, in which a false negative is fatal. Here, both are equally wrong.

In contrast, the evolution of the accuracy over time (See Fig. 12) indicates that initial epochs failed at generalizing. Later, the model started showing signs of overfitting, so epochs were configured with early stopping.

In addition, the use of a separate test set, that did not participate during the training process, provided an unbiased measure of 77% of accuracy, indicating the proportion of correct classifications.

As mentioned, precision and recall were also estimated, but the same information is conveyed, in a much more friendly way, by the confusion matrix (See Fig. 13): There is approximately 9% of false positives, and another 13% of false negatives.

## 5.3. Implications

Cross-validation was discarded due to performance issues. The computational resources required to train and evaluate the model on different folds of the data imposed a problematic constraint to the learning process. Future research is encouraged on a more powerful computer, with a GPU.

Other future improvements involve more advanced neural network architectures, such as transformers [11], and capturing more signals in the LHC. Obviously, doing that involves the participation of more experts in particle physics.

This model can be deployed in real-time systems for automated decision making regarding the presence of the Higgs Boson, contributing significantly to ongoing research in particle physics and potentially leading to new discoveries, and technological breakthroughs.

Interestingly, when the threshold, which decides whether the model confirms the presence of the Higgs boson, is reduced from 0.5 to 0.05, precision rises to 97%. Consequently, even if the model rarely confirms its presence, when it does, then we can be quite sure that we are in the presence of this exotic particule. As a result, this threshold will be used for real-time decisions, when thousands of events are generated per second. That way, Higgs bosons can be isolated and directed through a different magnetic field, where their properties can be more properly studied.

## 6. References

- [1] CERN. (n.d.). The Higgs Boson. Home.CERN. <https://home.cern/science/physics/higgs-boson>.
- [2] Whiteson,Daniel. (2014). HIGGS. UCI Machine Learning Repository.  
<https://doi.org/10.24432/C5V312>.
- [3] sldsrt. (2020). Higgs Boson Detection. Kaggle. <https://kaggle.com/competitions/higgs-boson-detection>.
- [4] Aggarwal, C. C. (2016). Recommender Systems: The Textbook. Springer.
- [5] Selvin, S. (1995). Practical Biostatistical Methods.
- [6] Sutton, R. S., & Barto, A. G. (Second edition). Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning series).
- [7] Baldi, P., P. Sadowski, and D. Whiteson (2014). Searching for Exotic Particles in High-energy Physics with Deep Learning. Nature Communications 5. <https://www.nature.com/articles/ncomms5308>
- [8] Jurecková, J. (2019). Methodology in Robust and Nonparametric Statistics. CRC Press.
- [9] Géron, A. (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow. O'Reilly
- [10] Ng, A. Deep Learning Specialization. Coursera. <https://www.coursera.org/specializations/deep-learning>
- [11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin (2017). Attention is all you need.

## 7. Annex

	Mean	Std	Min	Max		Mean	Std	Min	Max
signal	0.543922	0.498067	0.000000	1.000000	signal	5.439223e-01	0.498067	0.000000	1.000000
lepton pT	0.989767	0.559807	0.274697	5.851532	lepton pT	-2.016273e-17	1.000000	-1.277351	8.684719
lepton eta	-0.000035	1.008226	-2.428158	2.428050	lepton eta	9.819514e-18	1.000000	-2.408313	2.408275
lepton phi	-0.000010	1.006376	-1.742508	1.743236	lepton phi	-1.376352e-17	1.000000	-1.731459	1.732201
missing energy magnitude	0.995501	0.593610	0.000237	6.109808	missing energy magnitude	1.347212e-15	1.000000	-1.676629	8.615599
missing energy phi	0.000202	1.006232	-1.743616	1.743257	missing energy phi	9.337632e-18	1.000000	-1.733018	1.732259
jet 1 pt	0.989880	0.469982	0.137502	4.926362	jet 1 pt	-1.761214e-16	1.000000	-1.813639	8.375816
jet 1 eta	0.000014	1.008626	-2.961803	2.961752	jet 1 eta	1.537244e-17	1.000000	-2.936487	2.936408
jet 1 phi	0.000084	1.005877	-1.741237	1.741454	jet 1 phi	-3.459995e-17	1.000000	-1.731147	1.731195
jet 1 b-tag	0.999737	1.027822	0.000000	2.173076	jet 1 b-tag	-1.583309e-17	1.000000	-0.972675	1.141579
jet 2 pt	0.991545	0.495126	0.188981	5.573496	jet 2 pt	-1.539385e-16	1.000000	-1.620928	9.254109
jet 2 eta	-0.000161	1.008419	-2.909204	2.909324	jet 2 eta	-2.415120e-17	1.000000	-2.884757	2.885195
jet 2 phi	-0.000101	1.006131	-1.742372	1.743175	jet 2 phi	-9.988107e-18	1.000000	-1.731653	1.732653
jet 2 b-tag	0.998620	1.049301	0.000000	2.214872	jet 2 b-tag	-5.168752e-17	1.000000	-0.951700	1.159107
jet 3 pt	0.991414	0.483898	0.263608	5.525750	jet 3 pt	-1.620890e-16	1.000000	-1.504049	9.370434
jet 3 eta	-0.000091	1.007976	-2.727842	2.727278	jet 3 eta	-9.767742e-18	1.000000	-2.706168	2.705788
jet 3 phi	-0.000035	1.006320	-1.742069	1.742884	jet 3 phi	2.649025e-17	1.000000	-1.731094	1.731973
jet 3 b-tag	0.999263	1.193519	0.000000	2.548224	jet 3 b-tag	5.512508e-17	1.000000	-0.837240	1.297810
jet 4 pt	0.985864	0.502963	0.365354	5.433091	jet 4 pt	-1.263594e-16	1.000000	-1.233707	8.842052
jet 4 eta	-0.000090	1.006859	-2.496432	2.496343	jet 4 eta	-5.705593e-18	1.000000	-2.479336	2.479426
jet 4 phi	0.000024	1.006416	-1.742691	1.743372	jet 4 phi	6.019547e-18	1.000000	-1.731605	1.732234
jet 4 b-tag	1.000507	1.400379	0.000000	3.101961	jet 4 b-tag	5.177846e-17	1.000000	-0.714454	1.500633
m_jj	1.032058	0.652120	0.075070	13.583668	m_jj	-2.167807e-18	1.000000	-1.467501	19.247378
m_jjj	1.023499	0.370092	0.198676	7.271092	m_jjj	2.055753e-16	1.000000	-2.228695	16.881175
m_lv	1.050284	0.161760	0.083049	3.431373	m_lv	-1.135371e-16	1.000000	-5.979446	14.719879
m_jlv	1.007871	0.388983	0.132006	5.877596	m_jlv	3.980308e-17	1.000000	-2.251676	12.519111

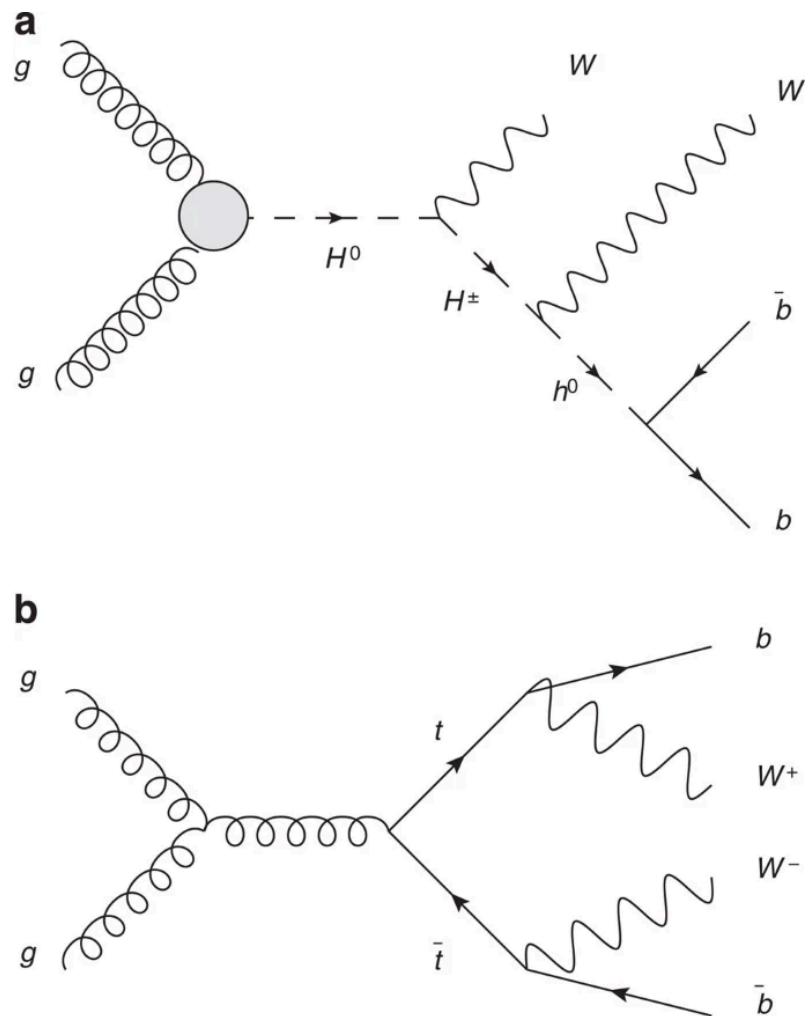
**Fig. 1:** Statistical summary showing the mean, standard deviation, minimum, and maximum values for each feature, before and after normalization.

signal	0	signal	0
lepton pT	1784	lepton pT	1073
lepton eta	0	lepton eta	1058
lepton phi	0	lepton phi	0
missing energy magnitude	1595	missing energy magnitude	1073
missing energy phi	0	missing energy phi	1063
jet 1 pt	1333	jet 1 pt	1073
jet 1 eta	0	jet 1 eta	1050
jet 1 phi	0	jet 1 phi	0
jet 1 b-tag	0	jet 1 b-tag	0
jet 2 pt	2409	jet 2 pt	1073
jet 2 eta	0	jet 2 eta	1021
jet 2 phi	0	jet 2 phi	0
jet 2 b-tag	0	jet 2 b-tag	0
jet 3 pt	2635	jet 3 pt	1073
jet 3 eta	0	jet 3 eta	991
jet 3 phi	0	jet 3 phi	0
jet 3 b-tag	0	jet 3 b-tag	0
jet 4 pt	1915	jet 4 pt	1073
jet 4 eta	0	jet 4 eta	987
jet 4 phi	0	jet 4 phi	0
jet 4 b-tag	0	jet 4 b-tag	0
m_jj	23449	m_jj	1073
m_jjj	17960	m_jjj	1073
m_lv	15054	m_lv	1073
m_jlv	8086	m_jlv	1073
m_bb	6250	m_bb	1073
m_wbb	6800	m_wbb	1073
m_wwbb	5589	m_wwbb	1073
dtype: int64		dtype: int64	

**Fig. 2:** Statistical summary illustrating the number of records with values exceeding 8 standard deviations from the mean (displayed on the left) and those exceeding the 99.99th percentile (displayed on the right).

feature		Feature	
lepton pt	0.093147	lepton pT	0.062
lepton eta	0.097098	lepton eta	0.067
lepton phi	0.097642	lepton phi	0.066
missing energy magnitude	0.088088	missing energy magnitude	0.060
missing energy phi	0.098752	missing energy phi	0.067
jet 1 pt	0.083687	jet 1 pt	0.056
jet 1 eta	0.097592	jet 1 eta	0.067
jet 1 phi	0.097753	jet 1 phi	0.066
jet 1 b-tag	0.098031	jet 1 b-tag	0.065
jet 2 pt	0.096489	jet 2 pt	0.066
jet 2 eta	0.097799	jet 2 eta	0.067
jet 2 phi	0.098614	jet 2 phi	0.065
jet 2 b-tag	0.095865	jet 2 b-tag	0.062
jet 3 pt	0.097414	jet 3 pt	0.067
jet 3 eta	0.098527	jet 3 eta	0.066
jet 3 phi	0.098207	jet 3 phi	0.066
jet 3 b-tag	0.098101	jet 3 b-tag	0.063
jet 4 pt	0.096810	jet 4 pt	0.067
jet 4 eta	0.096633	jet 4 eta	0.067
jet 4 phi	0.097816	jet 4 phi	0.067
jet 4 b-tag	0.095651	jet 4 b-tag	0.066
jj	0.098300	m_jj	0.067
jjj	0.095254	m_jjj	0.059
lv	0.095645	m_lv	0.065
jlv	0.094985	m_jlv	0.060
bb	0.066597	m_wbb	0.067
wbb	0.089777	m_wwbb	0.047
wwbb	0.074018		
Worst predictor is "m_bb"		Worst predictor is "m_wwbb"	

**Fig. 3:** Left image displays R-squared scores for all features excluding one, identifying 'm\_bb' as the least impactful. The right image shows recalculated scores after 'm\_bb' removal, with 'm\_wwbb' then identified as the next feature to remove for minimal R-squared reduction.



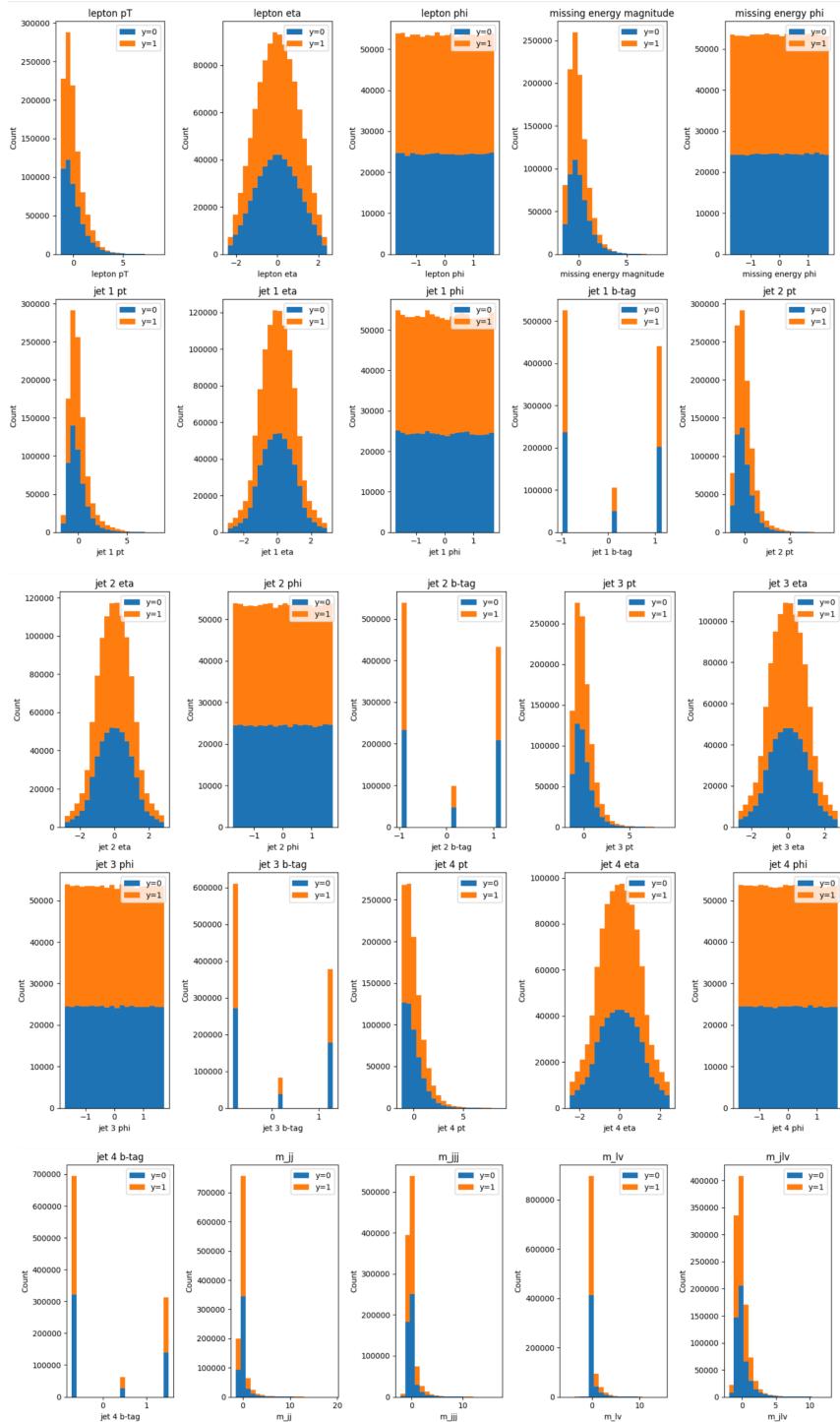
**Fig. 4:** (a) Diagram describing the signal process involving new exotic Higgs bosons  $H^0$  and  $H^\pm$ . (b) Diagram describing the background process involving top quarks ( $t$ ). In both cases, the resulting particles are two  $W$  bosons and two  $b$ -quarks. Source: [7]

	max	min	10%	20%	30%	40%	50%	60%	70%	80%	90%
<b>lepton pT</b>	8.684719	-1.277351	-0.997839	-0.802671	-0.622868	-0.440450	-0.244301	-0.019711	0.259147	0.658964	1.333062
<b>lepton eta</b>	2.408275	-2.408313	-1.340859	-0.907115	-0.572871	-0.278234	-0.000019	0.278195	0.571866	0.906110	1.341786
<b>lepton phi</b>	1.732201	-1.731459	-1.385765	-1.039520	-0.692723	-0.346476	-0.000230	0.346117	0.692914	1.039159	1.385956
<b>missing energy magnitude</b>	8.615599	-1.676629	-1.082155	-0.819424	-0.599072	-0.390506	-0.177399	0.055410	0.330454	0.693067	1.281261
<b>missing energy phi</b>	1.732259	-1.733018	-1.385949	-1.039112	-0.692818	-0.346235	0.000152	0.346575	0.692553	1.039054	1.385400
<b>jet 1 pt</b>	8.375816	-1.813639	-0.991481	-0.759336	-0.566368	-0.385486	-0.201485	-0.000915	0.237467	0.563758	1.183592
<b>jet 1 eta</b>	2.936408	-2.936487	-1.246876	-0.841409	-0.537063	-0.257261	-0.000039	0.256200	0.536984	0.841330	1.246797
<b>jet 1 phi</b>	1.731195	-1.731147	-1.388891	-1.037818	-0.691153	-0.349446	0.000077	0.349497	0.691201	1.037866	1.388388
<b>jet 1 b-tag</b>	1.141579	-0.972675	-0.972675	-0.972675	-0.972675	-0.972675	0.084452	1.141579	1.141579	1.141579	1.141579
<b>jet 2 pt</b>	9.254109	-1.620928	-0.998437	-0.773902	-0.583697	-0.398322	-0.204811	0.010569	0.267652	0.612463	1.200626
<b>jet 2 eta</b>	2.885195	-2.884757	-1.254914	-0.845527	-0.539209	-0.261789	0.000219	0.261264	0.538684	0.845965	1.254389
<b>jet 2 phi</b>	1.732653	-1.731653	-1.386446	-1.039033	-0.691621	-0.346962	0.000450	0.346862	0.692069	1.038379	1.386343
<b>jet 2 b-tag</b>	1.159107	-0.951700	-0.951700	-0.951700	-0.951700	-0.951700	-0.951700	1.159107	1.159107	1.159107	1.159107
<b>jet 3 pt</b>	9.370434	-1.504049	-0.1049660	-0.809037	-0.602893	-0.402918	-0.194959	0.035502	0.310240	0.672807	1.265110
<b>jet 3 eta</b>	2.705788	-2.706168	-1.264906	-0.854923	-0.545178	-0.266137	0.000262	0.265757	0.545702	0.855447	1.264526
<b>jet 3 phi</b>	1.731973	-1.731094	-1.385459	-1.039273	-0.692536	-0.347450	-0.000713	0.346678	0.692312	1.039601	1.385787
<b>jet 3 b-tag</b>	1.297810	-0.837240	-0.837240	-0.837240	-0.837240	-0.837240	-0.837240	0.230285	1.297810	1.297810	1.297810
<b>jet 4 pt</b>	8.842052	-1.233707	-1.017867	-0.825257	-0.637486	-0.443423	-0.233391	0.004711	0.292659	0.676428	1.302171
<b>jet 4 eta</b>	2.479426	-2.479336	-1.285759	-0.869703	-0.557867	-0.274155	0.000459	0.274245	0.557958	0.869793	1.285849
<b>jet 4 phi</b>	1.732234	-1.731605	-1.385342	-1.039630	-0.692815	-0.346550	-0.000287	0.346079	0.692893	1.039156	1.385971
<b>jet 4 b-tag</b>	1.500633	-0.714454	-0.714454	-0.714454	-0.714454	-0.714454	-0.714454	-0.714454	0.393089	1.500633	1.500633
<b>m_jj</b>	19.247378	-1.467501	-0.592056	-0.420924	-0.329867	-0.265405	-0.210225	-0.156817	-0.077798	0.095904	0.657581
<b>m_jjj</b>	16.881175	-2.228695	-0.760508	-0.555369	-0.414229	-0.300622	-0.196158	-0.085288	0.063958	0.289803	0.787000
<b>m_lv</b>	14.719879	-5.979446	-0.425828	-0.405540	-0.393544	-0.384971	-0.374045	-0.356749	-0.318345	0.112783	0.994902
<b>m_jlv</b>	12.519111	-2.251676	-0.926343	-0.709802	-0.536446	-0.385585	-0.234802	-0.053912	0.187743	0.540158	1.147213

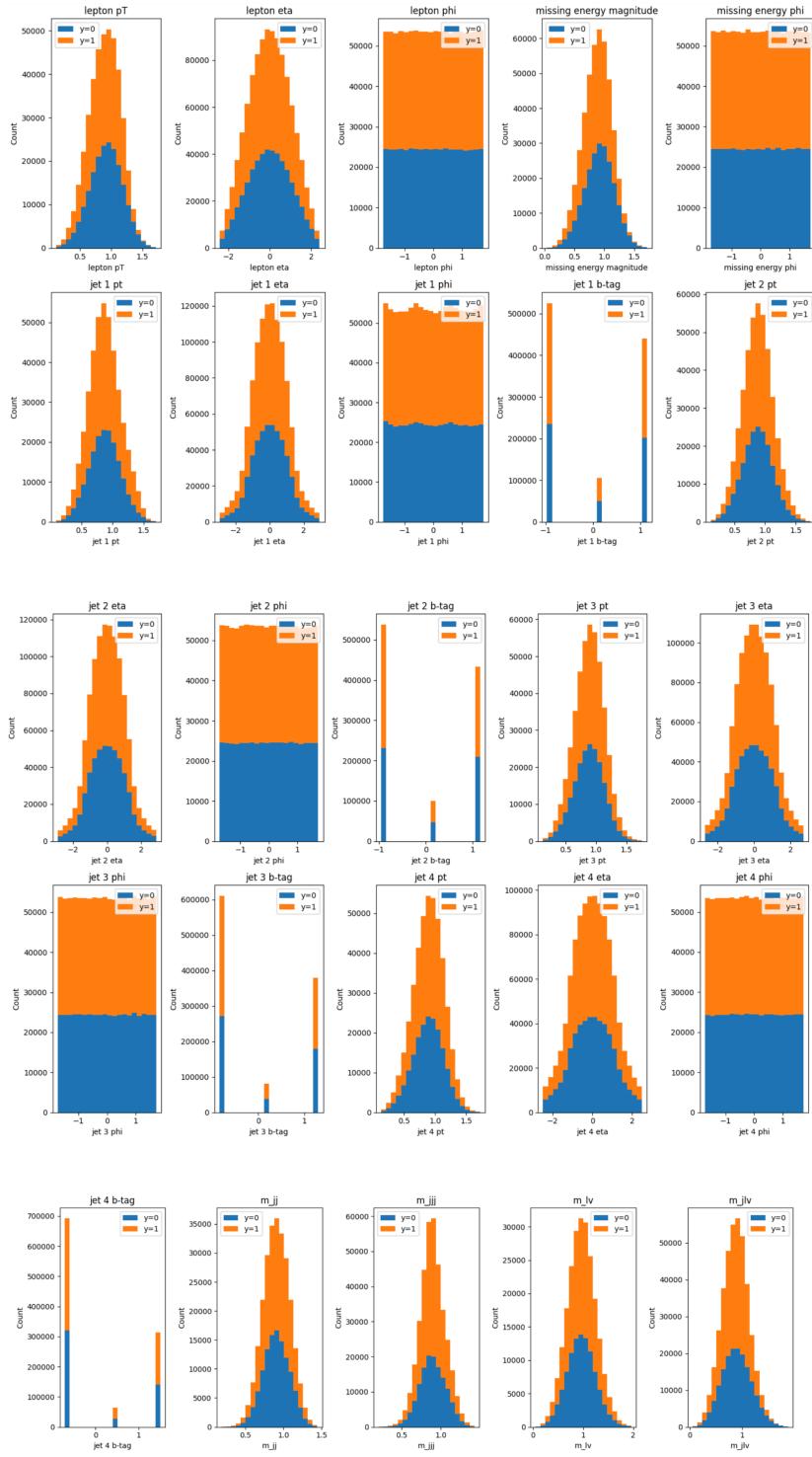
**Fig. 5:** Maxima, minima, and deciles from 10% to 90%.

	skew	kurt	median	iqr	trimmed_mean	winsorized_mean
lepton pT	1.662314	4.446112	-0.244301	1.149432	-0.137731	-4.096970e-02
lepton eta	-0.000086	-0.633241	-0.000019	1.464490	-0.000016	1.400163e-05
lepton phi	0.000428	-1.199795	-0.000230	1.731883	-0.000073	4.945425e-06
missing energy magnitude	1.389330	3.695907	-0.177399	1.202787	-0.106226	-3.689057e-02
missing energy phi	-0.000661	-1.199486	0.000152	1.731797	0.000101	6.967534e-06
jet 1 pt	1.803507	5.169810	-0.201485	1.044168	-0.134105	-4.641937e-02
jet 1 eta	0.000200	-0.015604	-0.000039	1.360721	-0.000018	-1.179883e-05
jet 1 phi	-0.000448	-1.200457	0.000077	1.726263	0.000047	1.019412e-05
jet 1 b-tag	0.159581	-1.863448	0.084452	2.114254	-0.021113	-1.583309e-17
jet 2 pt	1.845552	6.134492	-0.204811	1.100803	-0.127195	-4.688541e-02
jet 2 eta	0.000418	-0.053201	0.000219	1.376504	-0.000022	-5.902080e-05
jet 2 phi	0.000114	-1.199231	0.000450	1.729448	-0.000040	-1.997072e-06
jet 2 b-tag	0.196713	-1.857187	-0.951700	2.110807	-0.025926	-5.168752e-17
jet 3 pt	1.586724	4.870944	-0.194959	1.179887	-0.116219	-4.065963e-02
jet 3 eta	-0.000720	-0.196469	0.000262	1.387079	0.000119	3.271479e-05
jet 3 phi	0.000715	-1.200617	-0.000713	1.731585	-0.000082	-1.455435e-05
jet 3 b-tag	0.440632	-1.715070	-0.837240	2.135051	-0.057571	5.512508e-17
jet 4 pt	1.630674	4.418853	-0.233391	1.199702	-0.134758	-4.305734e-02
jet 4 eta	-0.000173	-0.395908	0.000459	1.416907	0.000037	-4.123661e-05
jet 4 phi	0.000250	-1.199747	-0.000287	1.731972	-0.000068	-2.824010e-05
jet 4 b-tag	0.757824	-1.343227	-0.714454	2.215088	-0.098272	5.177846e-17
m_ij	5.552663	49.161982	-0.210225	0.358661	-0.174510	-8.273098e-02
m_jjj	4.279232	31.546421	-0.196158	0.643401	-0.154376	-7.522295e-02
m_lv	4.190406	26.014672	-0.374045	0.215173	-0.242908	-7.286099e-02
m_jlv	2.516012	12.014491	-0.234802	0.964449	-0.141486	-5.560462e-02

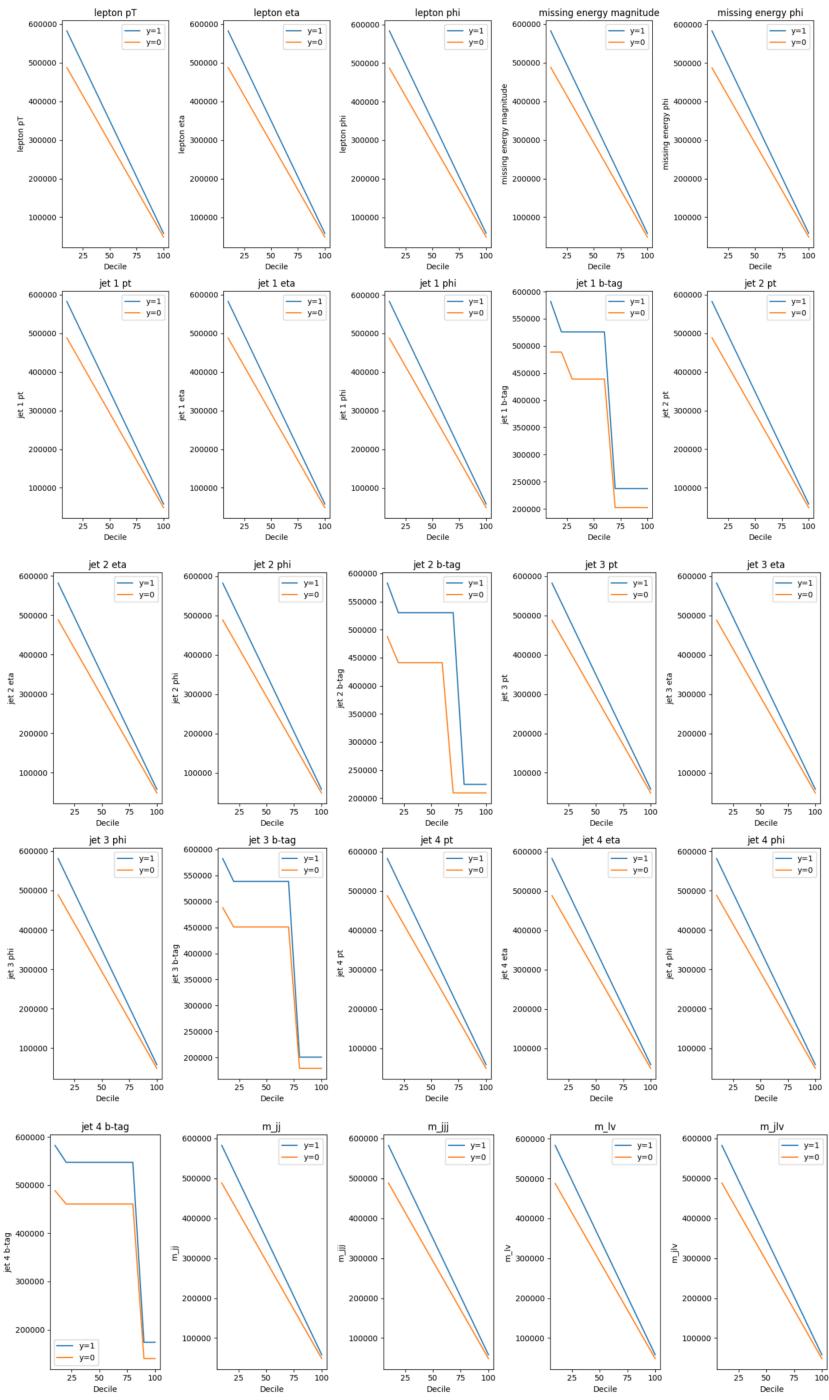
**Fig. 6:** Median, interquartile range, trimmed mean, and winsorized mean.



**Fig. 7:** Stacked histograms displaying the concentration of data points along the x-axis, with blue representing the absence of the Higgs boson and orange representing its presence.



**Fig. 8:** Stacked histograms after centering skewed features.



**Fig. 9:** Stacked line charts displaying the distribution of the deciles of the data points along the x-axis, with blue representing the presence of the Higgs boson and orange representing its absence.

Model: "sequential"

Layer (type)	Output Shape	Param #
<hr/>		
dense (Dense)	(None, 300)	7800
dense_1 (Dense)	(None, 300)	90300
dropout (Dropout)	(None, 300)	0
dense_2 (Dense)	(None, 300)	90300
dropout_1 (Dropout)	(None, 300)	0
dense_3 (Dense)	(None, 300)	90300
dropout_2 (Dropout)	(None, 300)	0
dense_4 (Dense)	(None, 300)	90300
dropout_3 (Dropout)	(None, 300)	0
dense_5 (Dense)	(None, 300)	90300
dropout_4 (Dropout)	(None, 300)	0
dense_6 (Dense)	(None, 300)	90300
dropout_5 (Dropout)	(None, 300)	0
dense_7 (Dense)	(None, 1)	301
<hr/>		
Total params: 549901 (2.10 MB)		
Trainable params: 549901 (2.10 MB)		
Non-trainable params: 0 (0.00 Byte)		

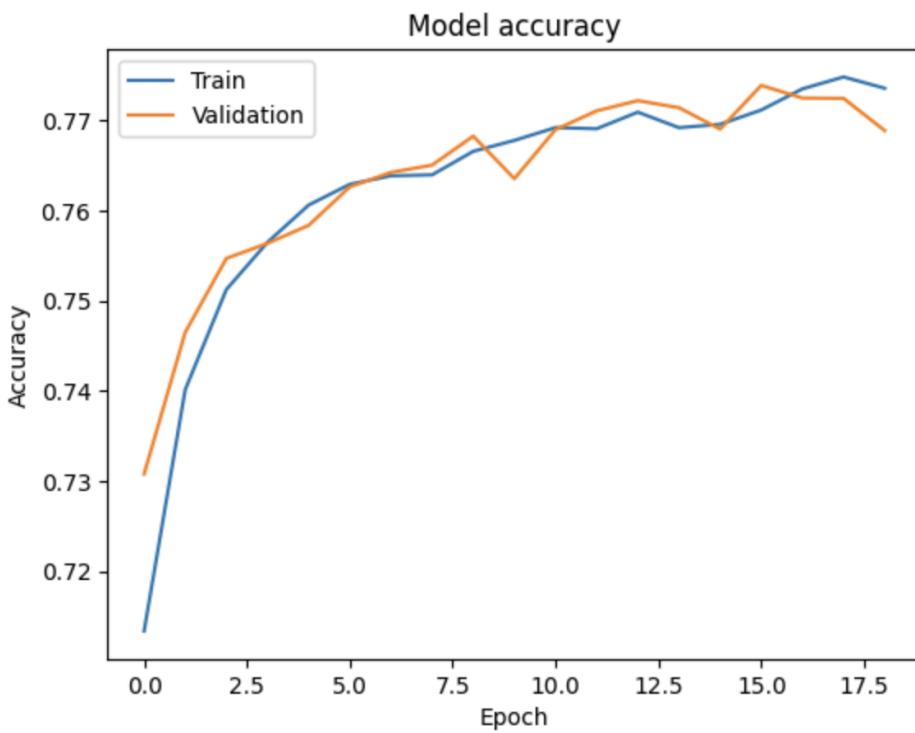
**Fig. 10:** Sequential model summary. To the right, the amount of parameters to optimize per hidden layer.

```

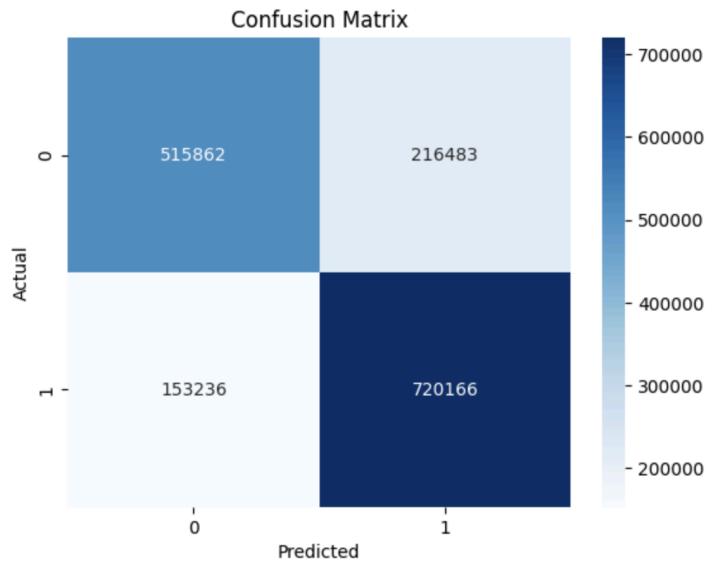
58543/58543 [=====] - 414s 7ms/step - loss: 0.4944 - accuracy: 0.7565 - precision: 0.7604 - recall: 0.8063 - val_loss
s: 0.4937 - val_accuracy: 0.7564 - val_precision: 0.7636 - val_recall: 0.7997
Epoch 5/25
58543/58543 [=====] - 399s 7ms/step - loss: 0.4876 - accuracy: 0.7606 - precision: 0.7639 - recall: 0.8102 - val_loss
s: 0.4879 - val_accuracy: 0.7584 - val_precision: 0.7819 - val_recall: 0.7707
Epoch 6/25
58543/58543 [=====] - 382s 7ms/step - loss: 0.4845 - accuracy: 0.7629 - precision: 0.7667 - recall: 0.8108 - val_loss
s: 0.4835 - val_accuracy: 0.7626 - val_precision: 0.7752 - val_recall: 0.7937
Epoch 7/25
58543/58543 [=====] - 376s 6ms/step - loss: 0.4825 - accuracy: 0.7639 - precision: 0.7659 - recall: 0.8150 - val_loss
s: 0.4821 - val_accuracy: 0.7642 - val_precision: 0.7635 - val_recall: 0.8208
Epoch 8/25
58543/58543 [=====] - 402s 7ms/step - loss: 0.4819 - accuracy: 0.7640 - precision: 0.7675 - recall: 0.8121 - val_loss
s: 0.4809 - val_accuracy: 0.7650 - val_precision: 0.7722 - val_recall: 0.8057
Epoch 9/25
58543/58543 [=====] - 424s 7ms/step - loss: 0.4782 - accuracy: 0.7666 - precision: 0.7696 - recall: 0.8147 - val_loss
s: 0.4752 - val_accuracy: 0.7683 - val_precision: 0.7752 - val_recall: 0.8083
Epoch 10/25
58543/58543 [=====] - 439s 7ms/step - loss: 0.4766 - accuracy: 0.7678 - precision: 0.7703 - recall: 0.8165 - val_loss
s: 0.4821 - val_accuracy: 0.7635 - val_precision: 0.7472 - val_recall: 0.8543
Epoch 11/25
58543/58543 [=====] - 463s 8ms/step - loss: 0.4747 - accuracy: 0.7692 - precision: 0.7715 - recall: 0.8179 - val_loss
s: 0.4747 - val_accuracy: 0.7690 - val_precision: 0.7784 - val_recall: 0.8042
Epoch 12/25
58543/58543 [=====] - 459s 8ms/step - loss: 0.4741 - accuracy: 0.7691 - precision: 0.7724 - recall: 0.8159 - val_loss
s: 0.4712 - val_accuracy: 0.7711 - val_precision: 0.7777 - val_recall: 0.8109
Epoch 13/25
58543/58543 [=====] - 455s 8ms/step - loss: 0.4716 - accuracy: 0.7709 - precision: 0.7736 - recall: 0.8183 - val_loss
s: 0.4692 - val_accuracy: 0.7722 - val_precision: 0.7651 - val_recall: 0.8386
Epoch 14/25
58543/58543 [=====] - 552s 9ms/step - loss: 0.4736 - accuracy: 0.7692 - precision: 0.7721 - recall: 0.8167 - val_loss
s: 0.4710 - val_accuracy: 0.7714 - val_precision: 0.7830 - val_recall: 0.8020
Epoch 15/25
58543/58543 [=====] - 466s 8ms/step - loss: 0.4735 - accuracy: 0.7696 - precision: 0.7717 - recall: 0.8185 - val_loss
s: 0.4741 - val_accuracy: 0.7690 - val_precision: 0.7629 - val_recall: 0.8349
Epoch 16/25
58543/58543 [=====] - 442s 8ms/step - loss: 0.4719 - accuracy: 0.7711 - precision: 0.7735 - recall: 0.8191 - val_loss
s: 0.4677 - val_accuracy: 0.7739 - val_precision: 0.7744 - val_recall: 0.8245
Epoch 17/25
58543/58543 [=====] - 488s 8ms/step - loss: 0.4680 - accuracy: 0.7735 - precision: 0.7757 - recall: 0.8209 - val_loss
s: 0.4713 - val_accuracy: 0.7725 - val_precision: 0.7854 - val_recall: 0.8005
Epoch 18/25
58543/58543 [=====] - 454s 8ms/step - loss: 0.4664 - accuracy: 0.7748 - precision: 0.7766 - recall: 0.8227 - val_loss
s: 0.4698 - val_accuracy: 0.7724 - val_precision: 0.7718 - val_recall: 0.8258
Epoch 19/25
58543/58543 [=====] - 454s 8ms/step - loss: 0.4682 - accuracy: 0.7736 - precision: 0.7759 - recall: 0.8209 - val_loss
s: 0.4761 - val_accuracy: 0.7689 - val_precision: 0.7675 - val_recall: 0.8250

```

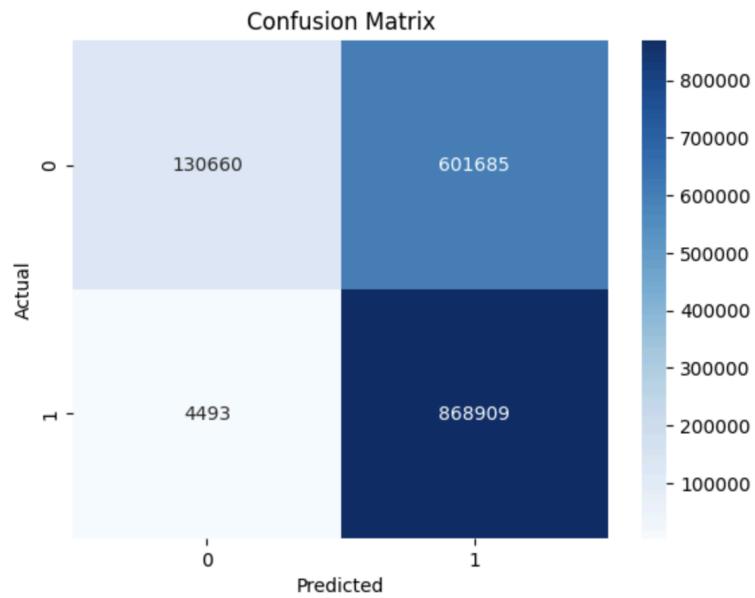
**Fig. 11:** History of the training step, showing the time per epoch, the loss and accuracy on the training set, and the loss and accuracy on the validation set.



**Fig. 12:** Line charts indicating how the accuracy of the model evolved over time, relative to the epochs. A blue line above the orange line suggests overfitting, and the opposite indicates poor generalization.



**Fig. 13:** A confusion matrix. True positives at the top left. False negatives at the top right. False positives at the bottom left. True positives at the bottom right.



**Fig. 14:** The confusion matrix when the threshold is reduced significantly, causing the precision to raise to 97%.