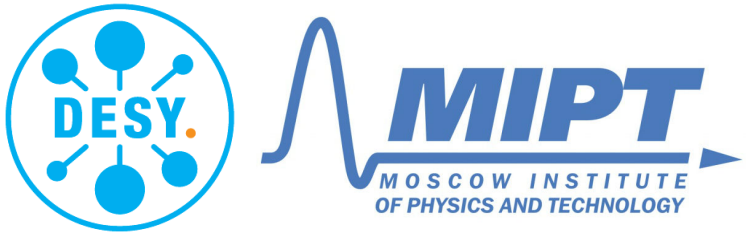


Search for $H^{\pm}$ boson with a deep neural network

Artem Nigoian, Moscow Institute of Physics and Technology, Russian Federation

Supervisors: Dirk Krucker, David Brunner

September 3, 2019

Abstract

In this research the solution of the charged Higgs boson four-momentum reconstruction problem in the process $qq' \to W^{\pm} \to hH^{\pm}$ (with further decays $H^{\pm} \to W^{\pm}h$, $W^{\pm} \to l\nu_l$ and both h decaying into $b\bar{b}$) in the THDM Type I model using deep neural network approach is considered. Different architectures of neural networks have been tested. Dense neural network turned out to be the most convenient one since it can be used to suppress $t\bar{t}$-background in addition to four-momentum reconstruction.

Contents

1 Introduction

The Standard Model of particle physics (SM) is the most successful theory that is able to make extremely precise predictions on the properties of the particles and the strength of their interactions. However, this theoretical model is not a complete description of the nature of elemantary particles as far as it does not explain Neutrino masses and their mixings, it has no candidate for Dark matter, etc. The most common way to solve these problems is to come up with an extension of the current theoretical model.
Such an extension can be looked for in the Higgs sector (very brief overview of the Higgs mechanism is given in the section 2.1). One of the scenarios is the so-called Two-Higgs-doublet model (also known as THDM). A particular specificataion of this model (it is described in details in the section) which provides us with an opportunity to insvestigate new regions in the space of parameters of this model is being considered in this research. Guided by the fact that a discovery of a new particle that can not be predicted within the SM is the most persuasive evidence of new physics, the process with intermediate h and $H^{\pm}$ is being studied. Since we want to test the given model (to figure out if it is in agreement with the experimental data or not), we should determine the final state as well as the intermediate particles (see section 2.3) and then be able to derive the kinematic properties of the intermediate particles using the information about the final state. To be more specific, we worked on the algorithm that will reconstruct four-momentum of $H^{\pm}$ for each event with the considered final state. We are mainly focused on Deep Neural Network (DNN) approach for the solution of the reconstruction problem (sections 3 and 4), but a brief discussion of another approach is also given in the very beginning of the mentioned section. The obtained results are given in section 5.

2 Physical constituent of the problem

2.1 Higgs mechanism and the «THDM»-extension

Higgs Mechanism is included into the SM to generate masses of the gauge bosons ($W^{\pm}$ and Z). This is achieved by introducing terms with a complex doublet Φ of scalar fields into the Electroweak Lagrangian $\mathcal{L}_{EW}$:

$$\mathcal{L}_{EW} = -\frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + (D_\mu\Phi)^\dagger(D_\mu\Phi) - \mu^2\Phi^\dagger\Phi - \lambda(\Phi\Phi^\dagger)^2, \quad (1)$$

where W^i_μ and B_μ are massless boson fields and D_μ is the covariant derivative, defined as

$$D_\mu = \partial_\mu - ig_2\frac{\tau_a}{2}W^a_\mu - ig_1\frac{Y_h}{2}B_\mu \quad (2)$$

After spontaneous symmetry breaking (see fig. 1) we will end up with three massive bosons – W^+, W^- (mixes W^1 and W^2) and Z (mix W^3 and B), – one massless boson γ (mix of W^3 and B) and massive CP-even neutral particle – Higgs boson.

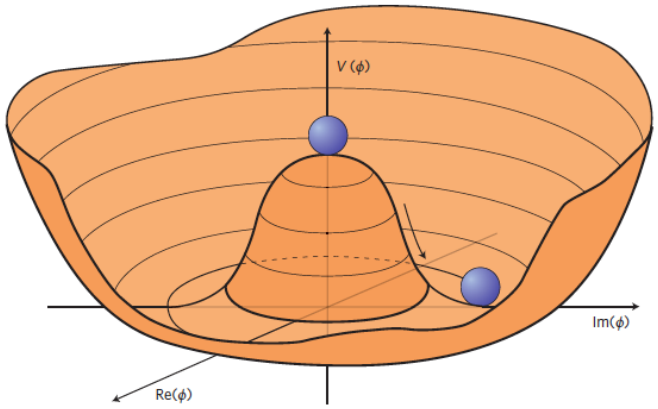


Figure 1: Higgs potential and symmetry breaking

However, nothing forbids the introduction of more Higgs-doublets. Actually, it is one of the ways to go beyond the Standard Model. In our case the simplest extension is considered – only one additional Higgs-doublet is introduced (now these doublets are commonly designated as Φ_1 and Φ_2) and since that the model is called the Two-Higgs-Doublet model (THDM). Having performed the symmetry breaking of the extended Lagrangian, we will end up with five Higgs-sector particles: two neutral CP-even Higgs bosons (h and H; H is heavier than h by convention), CP odd pseudoscalar particle A and two charged Higgs bosons ($H^{\pm}$).

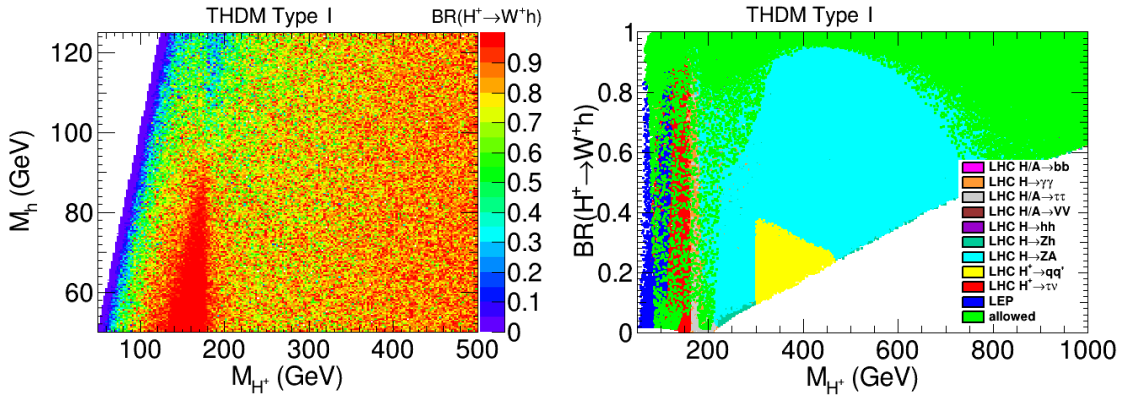
So, this model contains six free parameters: masses of $H^{\pm}$, h, H and A (4 parameters), angle α – the mixing angle which diagonalizes the mass matrix of the neutral CP even Higgs-particles, β – angle which is responsible for the ratio of the vacuum expectation values v_1 and v_2, corresponding to the doublets Φ_1 and Φ_2: $\tan\beta = v_1/v_2$.
It should be also mentioned that the model has been chosen not only due to the relative simplicity, but also due to a physical reason: it will be «natural» if one of the doublets couples to up-type quarks and the other – to down-type quarks.

2.2 Model specification

Even such a simple extension called 2HDM contains have six free parameters (see section 2.1). Some assumptions were made to clarify the model.
First of all, the extended Lagrangian is required to be symmetric under the transformation $\Phi_1 \to -\Phi_1$ to avoid Flavor-changing neutral currents. Then, $\sin(\alpha - \beta)$ is supposed to be equal to 0 (definitions of α and β are given in the section 2.1). This assumption leads to the fact that the discovered Higgs boson is heavier than the other CP even Higgs (since that H stands for the SM-Higgs boson and h – for non-SM one). Moreover, h does not couple to SM gauge bosons and the couplings to fermions are similar to those for the SM-Higgs H, so the b-decay and τ-leptons are dominant. Another very important implication is that $H^{\pm}$ decays only via $W^{\pm}$ and h (but not H). After that, the masses of A and $H^{\pm}$ are required to be equal to each other. And, finally, the angle β is fixed by $\tan\beta = 3$. At the end of the day, the specified model has only two free parameters ($m(H^{\pm})$ and $m(h)$) since four independent requirements on six initial parameters were imposed.
The branching ratio of the decay $H^{\pm} \to hW^{\pm}$ is quite high in our case (THDM Type I, fig. 2a) and, on the other hand, there is a large unexcluded region of $m(H^{\pm})$ and $BR(H^{\pm} \to hW^{\pm})$ parameters (see fig. 2b). Consequently, even though the model has been simplified significantly, the resulting one still provides us with the opportunity to obtain some useful results.

2.3 Considered process and topology

Having specified the model (see section 2.2), we should choose the process for studying and searching for in the data. The Feynman diagram of the process having been chosen is given on fig. 3.
This process has been chosen due to a number of reasons. As it was mentioned in the section 2.2 h strongly couples to b-quarks and does not couple to gauge bosons – this fact explains the consideration of the decay of h_1 and h_2 into $b\bar{b}$. Then, because $H^{\pm}$ decays to $W^{\pm}h$ (not $W^{\pm}H$) some background events some background events can be excluded from consideration. And, finally, the leptonic



a) b)

Figure 2: a) Branching ratio of the decay $H^+ \to hW^+$ as a function of $m(H^+)$ and $m(h)$ for THDM Type I. b) Excluded regions in the space of $m(H^+)$ and $BR(H^+ \to hW^+)$ variables.

decay of the final $W^\pm$ was chosen because otherwise the final state will contain six jets instead of four of them and the analysis will become more tricky.

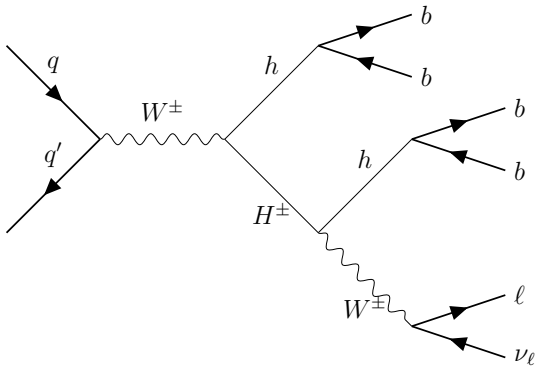


Figure 3: Feynman of the process being studied

On fig. 4 a typical event kinematic of this process in x-y plane (z axis is oriented along the beam momentum) in the so-called «non-boosted topology» (when final state contains exactly four non-overlapping b-jets) is shown. Only two of four jets origin from a charged Higgs, and one has to choose these jets correctly to reconstruct the four-momenta of the intermediate particle $H^\pm$ (or, at least, the mass of this particle). Two possible solutions of this combinatorical problem will be discussed in the next section.

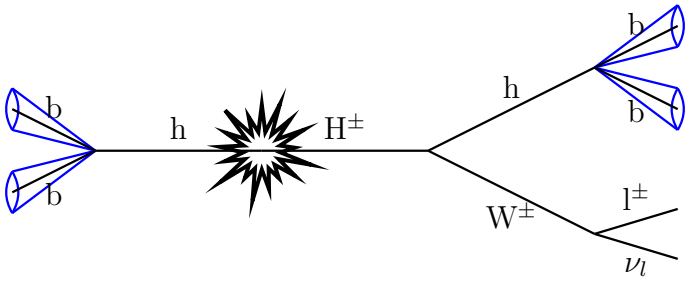


Figure 4: Typical event kinematic of the process in the x-y plane

2.4 Background discussion

While dealing with rare events, all possible background events that can be occasinally accepted as a signal should be taken into account. The most contributing background process in our case is $qq' \to t\bar{t}$ with further decays $t \to W^+b$ and $\bar{t} \to W^-\bar{b}$, where one W decays hadronically and the other – leptonically. Products of this process are 4 jets, 1 charged lepton and «missed» neutrino, i.e., the same final state as in the considered process, and, unfortunatelly, that makes this process indistinguishable from that being studied. The Feynman diagram of this process is given on fig. 5

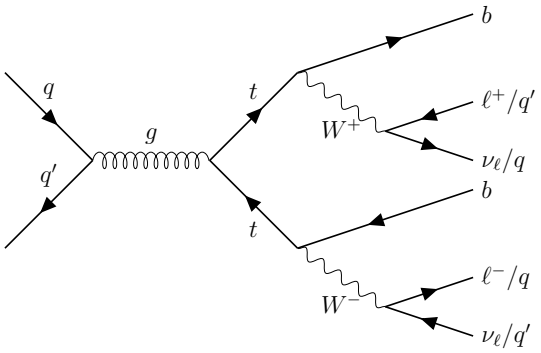


Figure 5: Feynman diagram of the most contributing background process.

3 Dense fully-connected neural network for $m(H^{\pm})$ reconstruction

3.1 Brief description of $H^{\pm}$ reconstruction approaches

There is a «physical» solution of the mass reconstruction problem. Since four jets b_1, b_2, b_3 and b_4 are given, two h-candidates (h_1 and h_2) can be built and then pair combination with minimal difference of their invariant masses should be chosen (both of h_1 and h_2 are the non-SM neutral Higgs-particles). After that one out of two $H^{\pm}$ candidates should be chosen with $max(\Delta\phi(H_i^{\pm}, h_j))$ and $min(\Delta\phi(H_i^{\pm}, h_i))$ (this choice is based on the two-body decay properties). Eventually, the mass of $H^{\pm}$ can be calculated.
Another solution that can be tested is one which is based on the DNN (Deep Neural Network) approach. As far as 4-momenta of all «particles» (here and below no difference is made between «jet» as «particle») are carrying all kinematic information about the event, these parameters are used as inputs of the network. And, clearly, the network should be able to produce four outputs with each of them corresponding to one of the four independent kinematic parameters of the charged Higgs-particle.
In this research only the DNN-based solution was consdidered.

3.2 Dense architecture

A dense fully-connected neural network that transforms all the measured kinematic variables of the final state into the kinematic variables of the «target-particle» $H^{\pm}$ was implemented. The graphical representation of the architecture of this network is given on fig. 6. Since data is slightly unbalanced (see more in section 3.3.1), weighted mean squared error as a loss function was used. To minimize the loss function Adam minimizer with initial learning rate 0.001 was chosen.

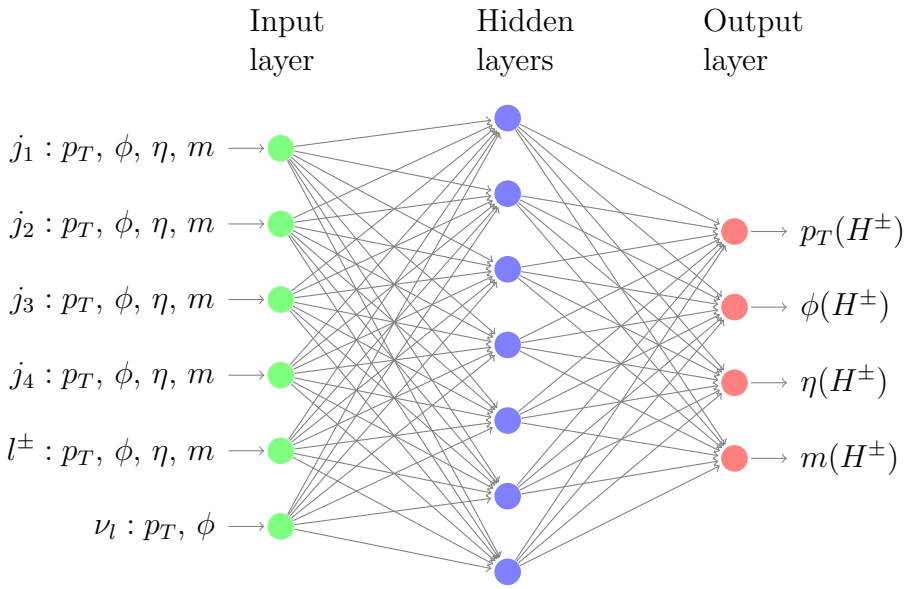


Figure 6: Dense fully-connected neural network with four outputs – kinematic parameters of $H^{\pm}$

3.2.1 Hyperparameters tuning

In case of the dense neural network number of layers, number of nodes in each layer, activation function and batch size are the parameters that should be defined to specify the network. To find the best set of these parameters random search in the four dimensional space of the parameters was implemented and used. The mentioned space is a Cartesian product of the following sets:

- number of layers (from 1 to 7 with step 1)
- number of nodes in each layer (from 50 to 250 with step 50)
- activation function (relu, elu, selu, softplus)
- batch size (from 25 to 250 with step 25)

The network with five layers each consisting of 200 nodes with softplus activation function and 100 samples in the batch turned out to have the best quality.

3.3 Data for training

3.3.1 Description of data used

To implement a solution based on DNN a training dataset is needed. As it was mentioned in the section 2.2, having specified the model we are still dealing with two free parameters which are masses of h and $H^{\pm}$. One of the parameters was fixed (by setting $m(h) = 100\ GeV$) and events for different masses of $H^{\pm}$ with two leptonic final states (corresponding to $W \to e\overline{\nu}_e$ – e-events – and $W \to \mu\overline{\nu}_\mu$ – μ-events) were generated with MadGraph (quark-level generation), hadronization was simulated with PYTHIA. After that all the events were passed through the Geant4 model of the CMS detector and, finally, they were reconstructed and since that used as a dataset for DNN. The distribution of the number of generated events is given in the table 1.

$m(H^{\pm})$, GeV	200	250	300	350	400	450	500	550	600
e-events	14274	16428	19778	22645	13553	12522	12717	7479	6678
μ-events	22332	18516	28679	28598	22303	18201	14342	10828	7850

Table 1: Distribution of the number of generated events with $m(h) = 100\ GeV$

Each event is represented by a vector of $4 \times 5 + 2 = 22$ features (4 kinematic parameters for each of the jets and charged lepton and 2 parameters for the missing lepton). Kinematic of each particle (except for neutrino) is described in terms of (p_T, ϕ, η, m), where p_T – transverse momentum, ϕ – angle between the $\vec{p}_T$ and x-axis, η – pseudorapidity ($\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$, where θ is the angle between the momentum of the particle and z-axis) and m – mass of the particle.
All the events were divided on training, test and validation in the following ratio: $81 : 10 : 9$.

3.3.2 Normalization of the inputs

As far as inputs with different scales (for instance, $\phi \in [-\pi, \pi]$ but m is not bounded) are being used, the inputs should be normalized. This point is crucial for Machine Learning algorithms but should not have a significant influence on the DNN-output. However, normalization of the inputs has no visible effect on the predictions, but leads to more stable training of the NN (see training curves on fig. 7). Here the scaler which centers each input to its' mean value and component wise scales to unit variance was used.

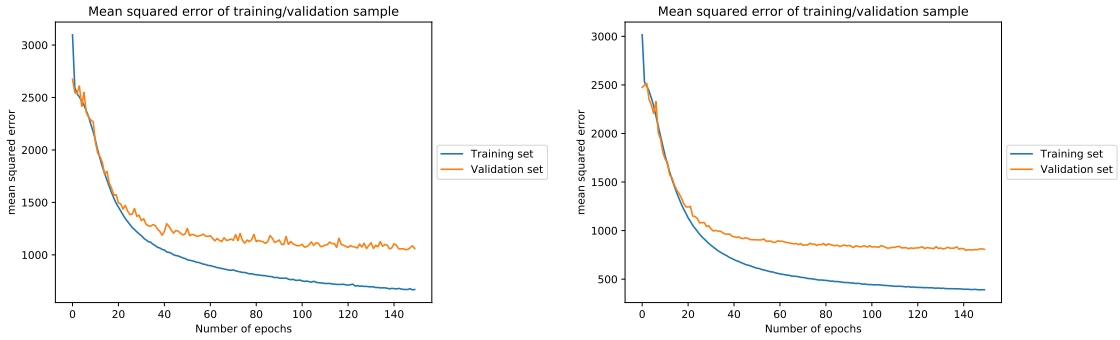


Figure 7: Comparison of training curves for the networks with same architectures but not normalized (left column) and normalized (right column) inputs.

3.3.3 Normalization of the outputs

Normalization of the outputs is also important – errors related to bounded parameters (like ϕ) have relatively small contribution to the loss function than the others (m or p_T). On fig. 8 predictions of the networks with same architectures but not normalized and normalized outputs are shown. Predicted distributions become being in agreement with data since all the outputs have comparable contributions.

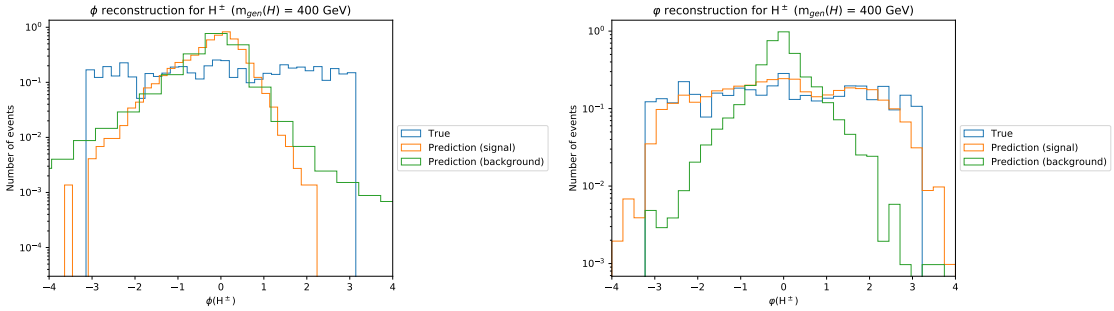


Figure 8: Comparison of the $\phi(H^{\pm})$ predictions of the networks with same architectures but not normalized (left column) and normalized (right column) outputs. Here and everywhere below blue histogram represents the real distribution of the events by mass; orange and green histograms represent the predicted distribution of the corresponding network for signal and background events respectively.

3.4 Number of input jets

On fig. 9 the distribution of the generated signal events by number of jets is given. The majority of the events ($\sim$ 80 %) contains exactly 4 jets, $\sim$ 18 % – 5 jets and the rest of the events has 6 or more jets. It should be mentioned that this distribution depends neither on the mass of $H^{\pm}$ nor on the type of lepton in the final state (e or μ).

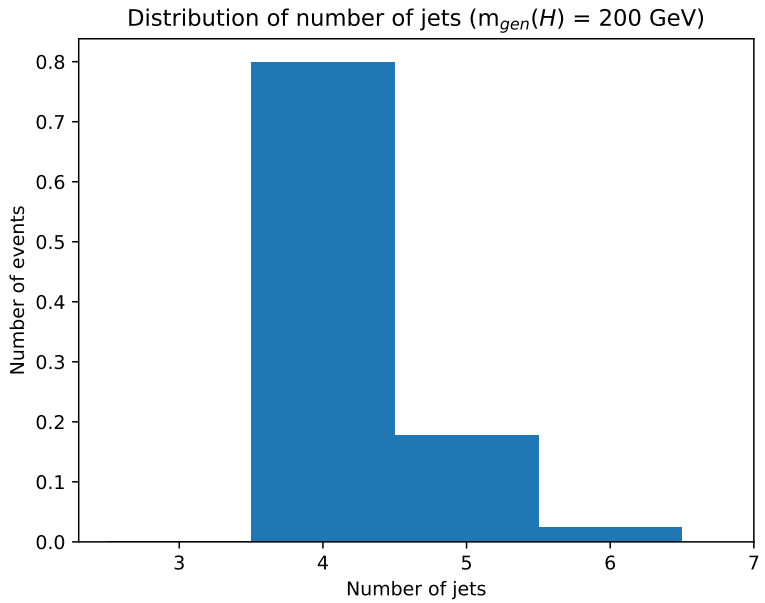


Figure 9: Distribution by number of jets of the generated signal events with $m(H^{\pm}) = 200\ GeV$

Due to this distribution it is quite clear that the improvement of the network's output cannot change dramatically if the information about the kinematic parameters of all the final jets (not only four of them with the maximal p_T) is included. However, this was tested and the results are given on fig. 10.
Due to barely noticeable change of the quality of the network 22 inputs (4 jets with highest p_T-s) are prefered.

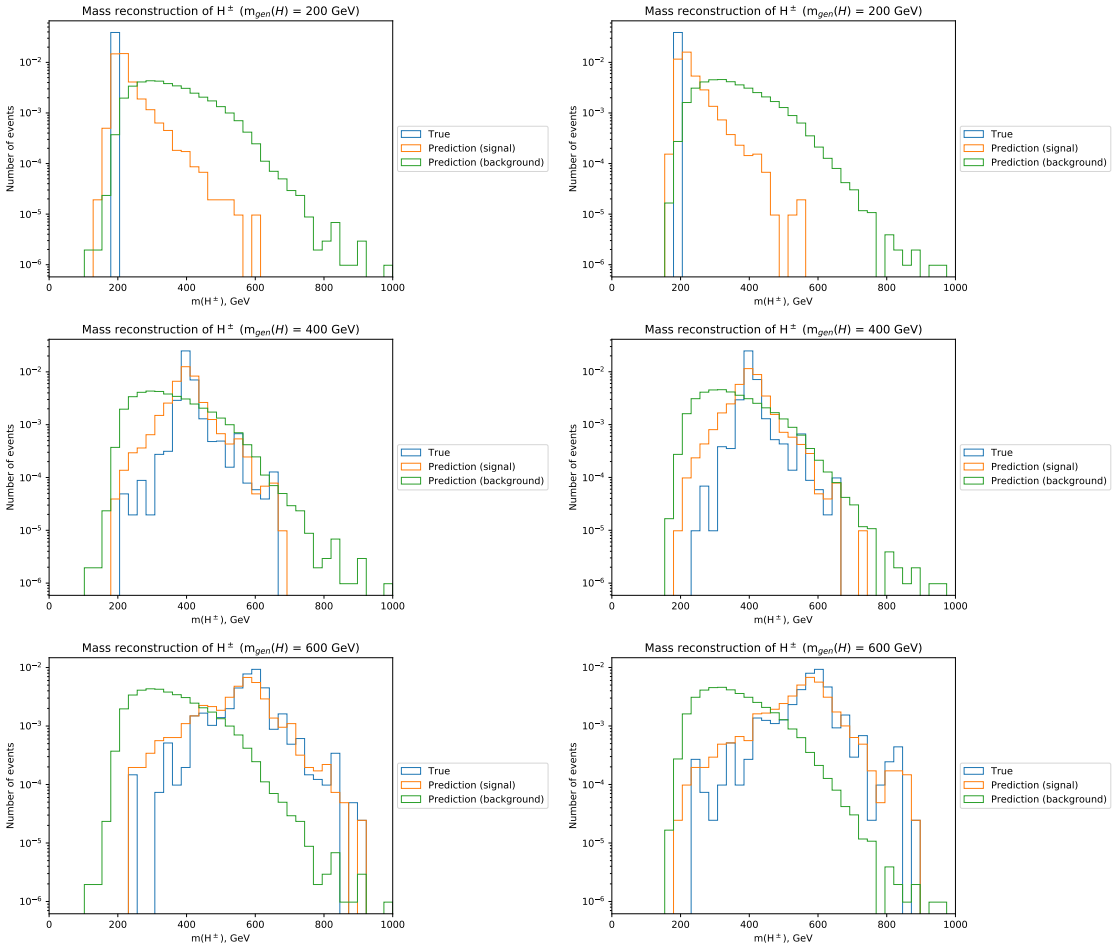


Figure 10: Comparison of the $m(H^{\pm})$ predictions of the networks with 4-jets (left column) and 6-jets (right column) inputs.

3.5 Convenient set of outputs

The network is aimed to reconstruct as much information about the intermediate $H^{\pm}$ as possible. To have a complete description of the kinematic properties of $H^{\pm}$ its' 4-momentum should be reconstructed. Possible candidates for the outputs are either (E, p_x, p_y, p_z) or (p_T, ϕ, η, m). The transformation from one set of variables to the other is simple, but both of them were tested as outputs to check if it affects the mass-reconstruction quality. It turned out to be preferable to have m in the outputs rather than derive it from predicted 4-momentum as $m = \sqrt{E^2 - \vec{p}^2}$ (see fig. 11) as the second method leads to wider distributions.

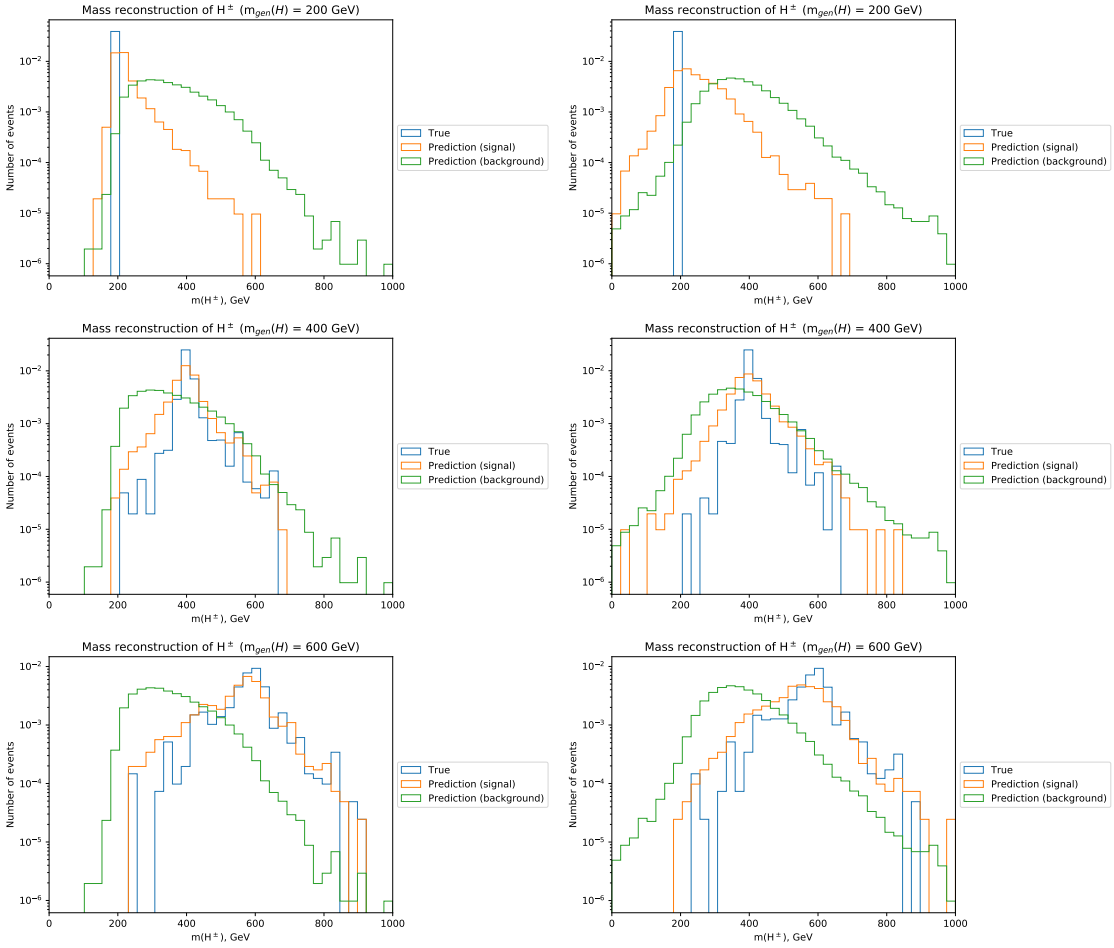


Figure 11: Comparison of the $m(H^{\pm})$ predictions of the networks with same architectures but (p_T, ϕ, η, m) (left column) and (E, p_x, p_y, p_z) (right column) outputs.

There is also another sufficient argument against neural network with (E, p_x, p_y, p_z) outputs (see section 3.6).

3.6 Background suppression

As it was mentioned in the section 2.4, the process being studied has a significant background $t\bar{t}$. To deal with this problem it was proposed to include background events into training and train the network such that it will predict for them vector $(p_T, \phi, \eta, m) = (0, 0, 0, 0)$. This network was trained, the outputs are given on fig. 12. Predicted distributions are similar to generated ones and, moreover, peak of the bakground is shifted to $m = 0\ GeV$ which provides us with an additional opprotunity to distinguish between signal and background events.

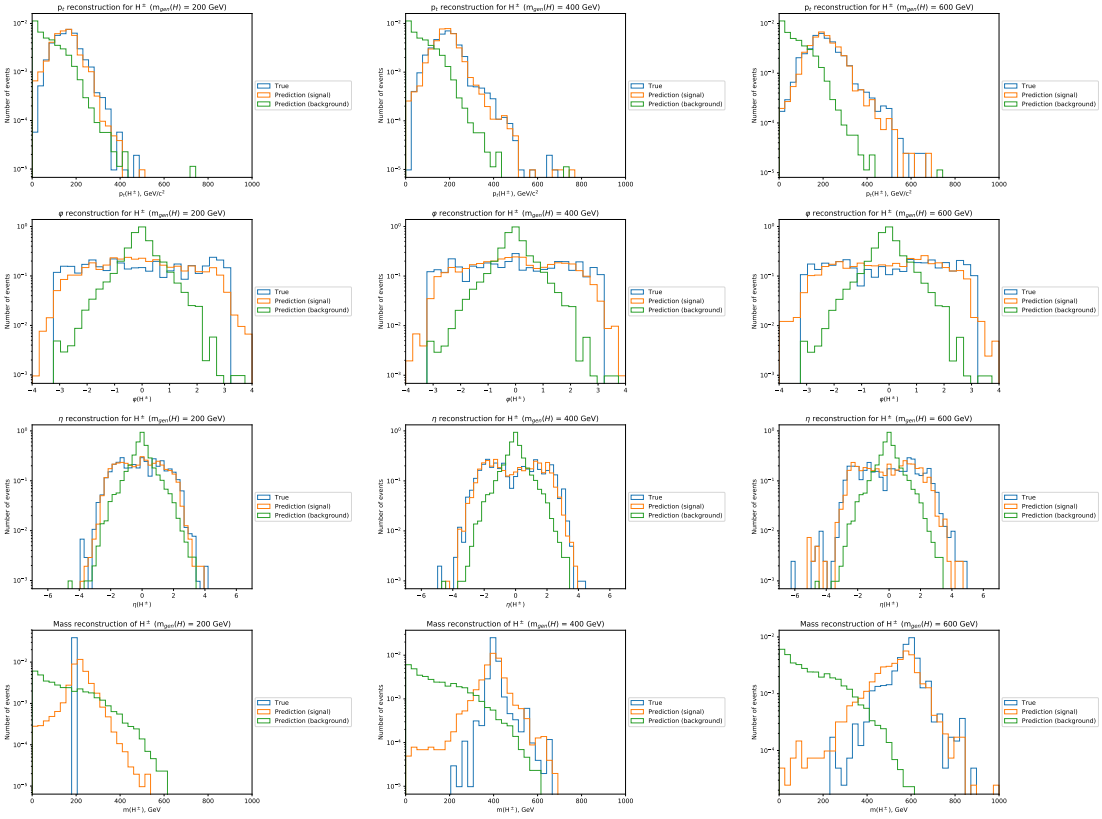


Figure 12: Predictions of the network with (p_T, ϕ, η, m) outputs (background is set to $(p_T, \phi, \eta, m) = (0, 0, 0, 0)$).

Training on the same data with (E, p_x, p_y, p_z) outputs was implemented. The obtained distributions (fig. 13) are shifted (in comparison to generated) and this is another argument against using of that set of outputs in the network.

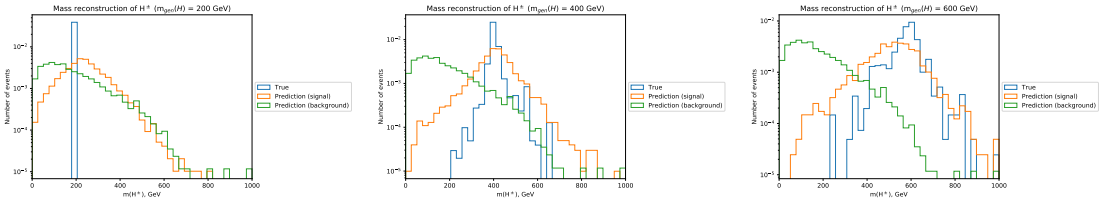


Figure 13: Predictions of the network with (E, p_x, p_y, p_z) outputs (background is set to $(E, p_x, p_y, p_z) = (0, 0, 0, 0)$).

3.7 Interpolation problem

Since the neural network was trained only on the events where masses of the $H^{\pm}$ were fixed in the range from 200 GeV to 600 GeV with 50 GeV step (see section 3.3.1), the prediction accuracy for intermediate masses should be estimated. To solve that problem, the network was retrained with data where $m(H^{\pm})$ was fixed in the range from 200 GeV to 600 GeV with 100 GeV step and then the obtained model was tested on events with $m(H^{\pm}) = 250$ GeV, $m(H^{\pm}) = 350$ GeV, $m(H^{\pm}) = 450$ GeV and $m(H^{\pm}) = 550$ GeV. Predictions are given on fig. 14. The conclusion is that generated mass can be reconstructed only if $m(H^{\pm})$ is relatively small. Therefore, to reconstruct precisely kinematical parameters of $H^{\pm}$ with $m(H^{\pm}) > 400$ GeV new generated events are needed.

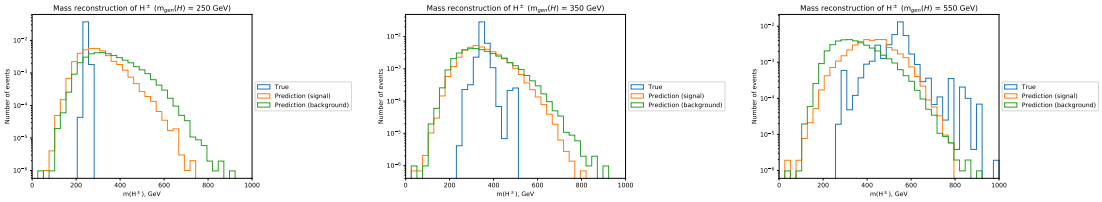


Figure 14: Predictions of the network trained on data with $m(H^{\pm}) = 200$, 300, 400, 500 and 600 GeV

4 Other architectures overview

Despite the fact that dense fully-connected neural network is able to give accurate predictions (see section 5), other architectures were tested.
Specifically, based on some positive results (see "ML applications to jet tagging in CMS", Mauro Verzetti) convolutional input layers before Dense Layers were added (see fig. 15). To implement this kind of neural network each input vector was reshaped into a table 6×4 (particle $\times$ momentum) where all particles were

ordered by p_t ((η, m) of neutrino were setted to $(0, 0)$). This approach turned out to be less accurate than the approach with Dense Layers.

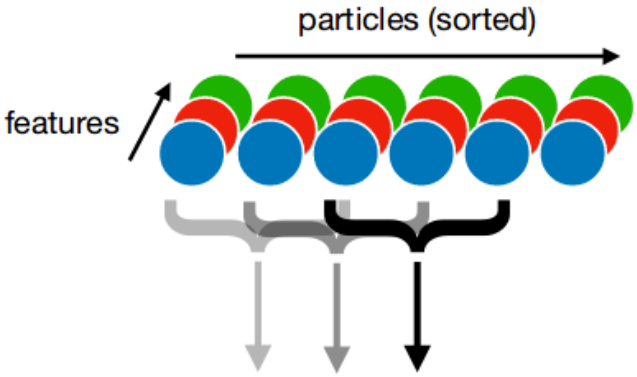


Figure 15: Graphical representation of the convolutional layer

As far as each event can be interpreted as a sequence of 4-momenta of the particles in the final state, recurrent neural network (RNN) as an input layer can be used. One of the most popular kinds of RNN called LSTM (long short term memory) was tested. Predictions of a modification of such a network are given on fig. 16 and they are in the agreement with generated distributions.

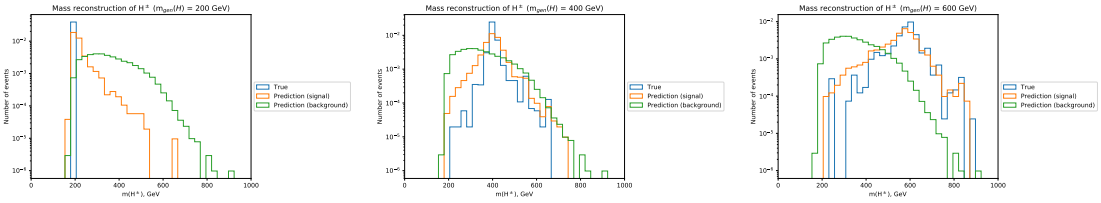


Figure 16: Predictions of $m(H^{\pm})$ obtained with the network with 1 LSTM input layer (128 cells) and 4 successive fully-connected layers.

Despite comparable quality of dense fully-connected network and one with LSTM-input layer, the first network is prefered for its' simplicity and less number of parameters. However, more detailed studies of other architectures are required.

5 Results

During this research application of the DNN approach for the charged Higgs boson four-momentum reconstruction problem in the given process has been considered. The method turned out to be successful. Dense neural network with five layers each consisting of 200 nodes with softplus activation function and 100 events in the batch has been trained on data without background events (fig. 17) and with them (fig. 18). The second network succeeds in distinguishing between signal and background events. Neural networks with LSTM and convolutional layers have been tested; however, none of them showed significantly better results.

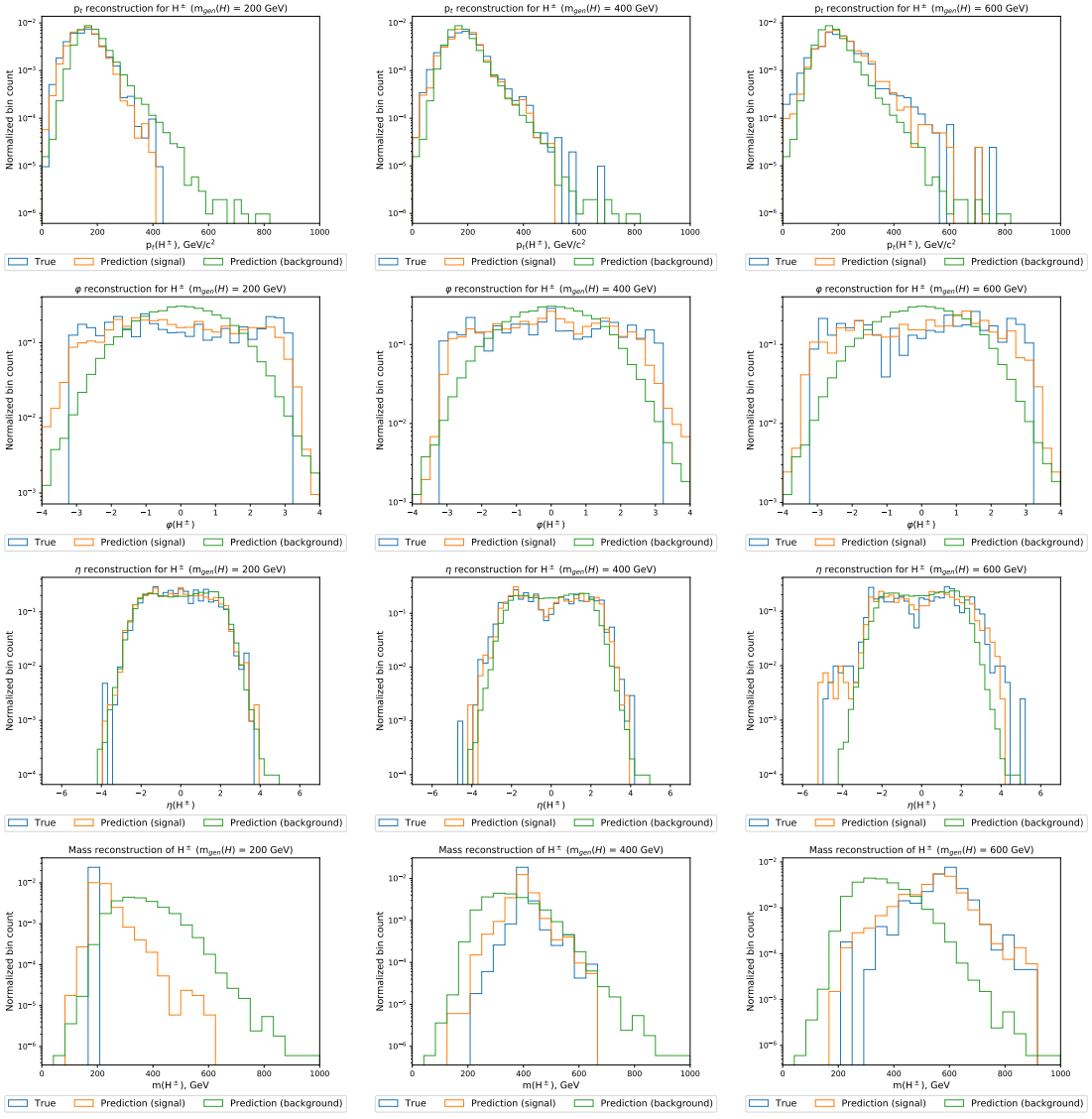


Figure 17: Predictions of the resulting dense fully-connected neural network (background events are not included into training)

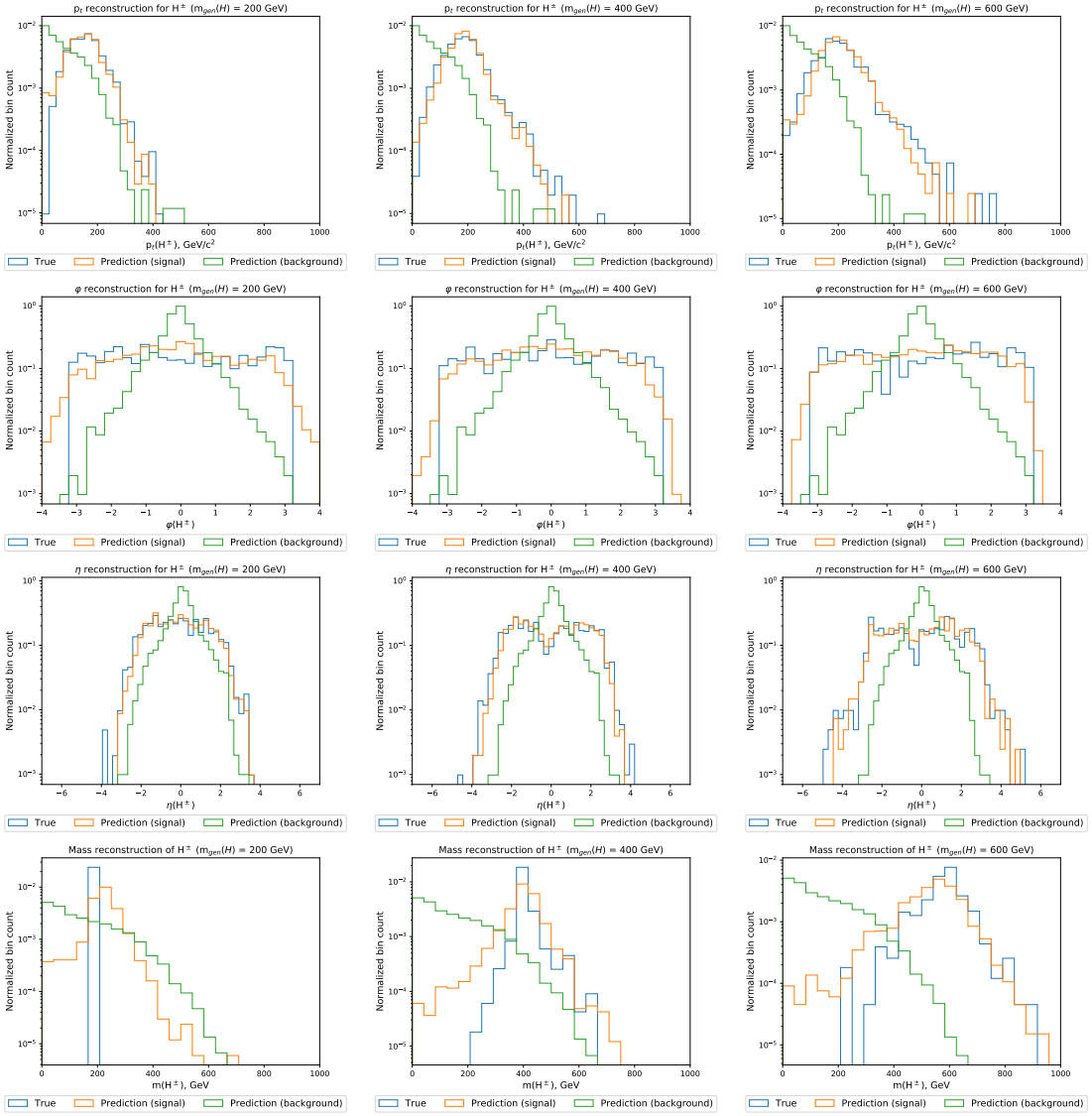


Figure 18: Predictions of the resulting dense fully-connected neural network (background events are included into training)

6 Acknowledgements

I want to thank organizing team of DESY Summer Student Program 2019 for constant informing, SUSY-group at DESY for regular discussion and helpful scientific advicement during these two months, and, of course, I would like to thank my supervisors Dirk Krucker and David Brunner for the provided opportunity to participate in this research.

References

[1] The Standard Model Higgs Boson *Part of the Lecture Particle Physics II, UvA Particle Physics Master 2013-2014*

[2] Theory and phenomenology of two-Higgs-doublet models *G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, Joao P. Silva*

[3] Introduction to Deep Learning with Keras *Lisa BENATO, Patrick L.S. CONNOR, Dirk KRUCKER, Mareike MEYER*

[4] The Most Intuitive and Easiest Guide for Recurrent Neural Network *Jiwon Jeong*

[5] Machine learning applications to jet tagging in CMS *Mauro Verzetti (CERN and FWO) on behalf of the CMS Collaboration*