Reconstruction of τ lepton pair invariant mass using an artificial neural network

P. Bärtschi [a],[*], C. Galloni [a], C. Lange [b], B. Kilminster [a]

[a] *Physik Institut, Universität Zürich, CH-8057 Zurich, Switzerland*
[b] *CERN, CH-1211 Geneva, Switzerland*

ARTICLE INFO

ABSTRACT

The reconstruction of the invariant mass of τ lepton pairs is important for analyses containing Higgs and Z bosons decaying to $\tau^+\tau^-$, but highly challenging due to the neutrinos from the τ lepton decays, which cannot be measured in the detector. In this paper, we demonstrate how artificial neural networks can be used to reconstruct the mass of a di-τ system and compare this procedure to an algorithm used by the CMS Collaboration for this purpose. We find that the neural network output shows a smaller bias and better resolution of the di-τ mass reconstruction and an improved discrimination between a Higgs boson signal and the Drell–Yan background with a much shorter computation time.

1. Introduction

Tau leptons can be produced singly or in pairs through the decay of heavier mesons and baryons, or through the decay of Standard Model bosons that are created in particle collisions. The Z^0/γ^* and Higgs (H) bosons are the only particles that can mediate resonant di-τ production, and Higgs boson production has been recently observed in this decay by the CMS and ATLAS experiment at CERN [1,2]. The reconstruction of the di-τ system invariant mass is fundamental in distinguishing between the Z and the Higgs boson. The dominant background in identifying the H $\rightarrow \tau^-\tau^+$ signal is the Drell–Yan (DY) process of di-τ production, therefore, it is important to distinguish between the two processes. However, the reconstruction of the mass of the di-τ system is challenging. The τ lepton decays after a short time into leptons or hadrons, both accompanied by neutrinos. Charged leptons and hadrons can be observed in the detector, thus they are called visible particles in this paper. Neutrinos, on the other hand, interact very weakly with matter and escape the experiment undetected, but can be identified in a relatively hermetic detector as used in the CMS and ATLAS experiments [3,4] as an imbalance in the measured momentum calculated in the transverse plane. The missing transverse momentum (MET) is defined as the negative vectorial sum of momenta in the transverse plane of all the visible particles produced in the collision. The momentum of the neutrinos in the beam direction, on the other hand, cannot be quantified, because it depends on the momentum of the quarks and gluons within the proton before the collision, which is only known statistically and not in any individual event.

The CMS Collaboration makes use of two different SVfit algorithms for the reconstruction of the di-τ mass. One is based on a likelihood method to reconstruct the mass on an event-by-event basis [5], while the second, improved algorithm employs a likelihood function of arbitrary normalization [6]. The latter algorithm is additionally able to reconstruct the kinematic properties of the di-τ system. The ATLAS and CDF Collaborations use a Missing Mass Calculation (MMC) method that is based on minimizing a likelihood function in the kinematically allowed detector phase space [7]. The method presented in this paper is based on an artificial neural network (called neural network or NN from now on) to reconstruct the mass of the di-τ system. The neural network is implemented with the Python deep learning library "Keras" [8] and can be trained with a data set of simulated di-τ events, which contain all the known parameters of the visible particles and missing energy from neutrinos as input, and the mass of the simulated di-τ system as a training target. After the training, the neural network is able to calculate an approximate value of the mass of the di-τ system using the known parameters of the visible particles and the missing energy for any event with two τ leptons.

The paper is structured as follows: in Section 2, events with simulated τ-lepton pairs are described, and are used to train and test the neural network. Section 3 focuses on the specific configuration of the neural network used to deliver the best predictions in comparison with the true mass of the simulated di-τ system. The performance of the neural network is compared to a standard tool for the di-τ mass reconstruction used by the CMS Collaboration in Section 4. Conclusions are presented in the final section.

---

* Corresponding author.
E-mail address: pascal.baertschi@uzh.ch (P. Bärtschi).

2. Simulation and selection of τ lepton pair events

Monte Carlo simulation is used to generate $\tau^+\tau^-$ events at different masses. The events are generated using `MadGraph_aMC@NLO` `2.5.5` [9], and then showered using `PYTHIA 8.226` [10,11]. The detector response is modeled using a simplified fast simulation (`DELPHES 3.4.1` [12]) of the phase-0 CMS detector with the acceptance and expected performance of the detector [3]. No additional proton–proton collisions (pile-up) are simulated.

DY events occur when a mediator Z^0/γ^* is produced through quark–antiquark annihilation. This mediator can then decay into a pair of τ leptons.

In proton collisions, the main production processes for the Higgs boson, ordered from largest cross section to smallest, are gluon–gluon fusion, vector boson fusion, W and Z associated production, and t associated production. Events are simulated using the most dominant production mode via gluon–gluon fusion, which occurs through an intermediate heavy-quark loop that is dominated by the top quark [13]. In order to increase the mass range over which the NN can reconstruct the di-τ system, the Higgs boson mass is artificially varied from 80 to 300 GeV in 5 GeV steps. The fully leptonic, semi-leptonic and fully hadronic Higgs decay channels are included.

Events with produced Higgs bosons must pass selection requirements similar to those applied in the CMS H $\rightarrow \tau^+\tau^-$ search [6]. The MET must be greater than 20 GeV for all events, in order to make sure that the neutrino momenta are not pointing in opposite directions. For events where both τ leptons decay leptonically, the e/μ with the higher p_T must satisfy $p_\mathrm{T} > 20$ GeV and $|\eta| < 2.4$, with the other e/μ satisfying $p_\mathrm{T} > 10$ GeV and $|\eta| < 2.4$. For events where only one τ lepton decays leptonically, the e/μ must satisfy $p_\mathrm{T} > 20$ GeV and $|\eta| < 2.1$, with the vectorial sum of the visible decay products of the hadronically decaying τ lepton having $p_\mathrm{T} > 30$ GeV and $|\eta| < 2.3$. For the case of events with two τ leptons decaying hadronically, both vectorial sums must have $p_\mathrm{T} > 20$ GeV and $|\eta| < 2.1$.

Fig. 1 shows the generated mass and visible mass spectrum for the simulated Higgs boson decaying to τ leptons. The mass corresponds to the true value of the mediator, while the visible mass is obtained by summing the momenta of the final state decay particles, not including the neutrinos. The di-τ_{gen} mass spectrum is not flat due to selection requirements that are related to detector acceptance and identification effects. The visible mass of the di-τ system (visible di-τ_{gen} mass) shows a much different spectrum. In order to separate signal from background, it is desirable to achieve a di-τ mass as close to the generated value as possible, and so information about the neutrino products must be inferred to improve the mass resolution. To prevent the neural network from learning the mass distribution of the training sample instead of the relationship between the inputs and the training target, the training sample consists of events from all the different Higgs boson masses, with the same number of events for each mass value.

The neural network is trained with 270'000 events and the performance of the neural network is tested with 100'000 independent events. The inputs for the neural network are the following parameters: 4 binary numbers classifying the decay channel (fully leptonic, semi-leptonic or fully hadronic), the p_T, η, ϕ, energy, and invariant mass of the visible decay products from each τ lepton, the MET and ϕ of the MET vector, and the collinear di-τ mass. The collinear di-τ mass is the mass of the di-τ system computed assuming that the vectorial sum of the visible τ lepton decay products point along the original τ lepton direction. The collinear di-τ mass can be calculated as shown in Eq. (1) [14]:

$$m_{\mathrm{coll}} = \frac{m_{di\text{-}\tau_{\mathrm{vis}}}}{\sqrt{\frac{p_{\mathrm{T}_{\tau_{\mathrm{vis1}}}}}{p_{\mathrm{T}_{\tau_{\mathrm{vis1}}}}+\mathrm{MET}_1} \cdot \frac{p_{\mathrm{T}_{\tau_{\mathrm{vis2}}}}}{p_{\mathrm{T}_{\tau_{\mathrm{vis2}}}}+\mathrm{MET}_2}}}, \qquad (1)$$

where m_{coll} stands for the collinear di-τ mass, $m_{di\text{-}\tau_{\mathrm{vis}}}$ for the visible di-τ mass, $p_{\mathrm{T}_{\tau_{\mathrm{vis1,2}}}}$ and $\mathrm{MET}_{1,2}$ for the visible transverse momentum and missing transverse momentum of τ_1 and τ_2, respectively.
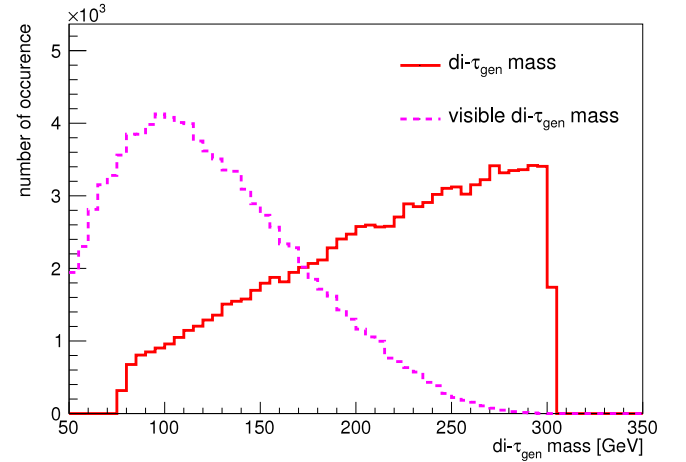

Fig. 1. Mass distribution of the generated events with applied cuts.

Table 1
Neural network model for reconstructing the Higgs boson mass from the di-τ system.

Layer	Number of neurons	Activation function
Input layer	17 neurons	–
1. Hidden layer	200 neurons	ReLU
2. Hidden layer	200 neurons	ReLU
3. Hidden layer	200 neurons	ReLU
4. Hidden layer	200 neurons	ReLU
Output layer	1 neuron	ReLU

3. Model of the neural network

The neural network used is feedforward, has fully connected layers, and is written with the Python deep learning library, Keras [8]. Keras is a set of high-level building blocks that implement the deep learning model, and it interfaces with a backend, which handles operations such as tensor products and convolutions. The backends used in this work are Theano [15] for running on CPU and Tensorflow [16] for running on GPU. The model used for reconstructing the Higgs boson mass from the di-τ system is shown in Table 1.

Batch sizes of 128 and 400 epochs are used for the training. The activation function used is the commonly employed "rectified linear unit (ReLU)" and the loss function is the mean squared error (MSE), which is described by Eq. (2):

$$C(w,b) = \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i(w_i, b_i) - Y_i)^2, \qquad (2)$$

where $\hat{Y}_i$ is the prediction, Y_i is the training target value, n is the number of the training events, w_i are the weights and b_i are the biases used for the prediction $\hat{Y}_i$. The optimizer used is called `adam` which stands for Adaptive Moment Estimation [17]. While the choice of network structure has been evaluated carefully and seems rather optimal for the problem and data set under investigation, the structure would have to be adjusted in case of adding or removing variables.

Once the neural network is trained, the performance can be tested with an independent second sample of simulated events. A good quantifier for the performance of the neural network is the mean squared error of the deviation between the predicted value and the true value of the mass from the simulated di-τ system of the test sample, as well as the reconstructed di-τ mass resolution.

4. Results

4.1. SVfit

The performance of the neural network is evaluated with respect to other current working methods for the reconstruction of the di-τ mass.
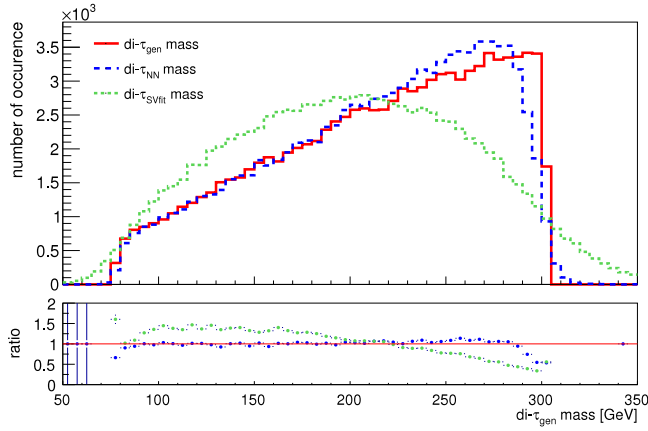
Fig. 2. Reconstructed mass of di-τ system. Di-τ_{gen} refers to the simulated mass values, di-τ_{NN} and di-τ_{SVfit} to the reconstructed mass values using the neural network and SVfit, respectively.

Fig. 3. The loss, as determined by the MSE, for the training and testing samples of the neural network.

For example the CMS Collaboration uses SVfit [5], which reconstructs the mass of the di-τ system using dynamical likelihood techniques. The term dynamical likelihood techniques refers to likelihood-based methods used for the reconstruction of kinematic quantities on an event-by-event basis. The inputs to SVfit are the visible decay products of the τ leptons, METx and METy as well as the MET covariance matrix. The MET covariance matrix represents the expected resolution of the MET reconstruction in the detector (since the MET measurement is affected by the accuracy of the energy calibrations).

4.2. Comparison between neural network and SVfit

If not specified differently, events with fully leptonic, semi-leptonic and fully hadronic decays are used. The predictions of the neural network are shown in Fig. 2 in comparison with the results from SVfit and with the mass of the simulated events (di-τ_{gen} mass). The shape of the di-τ_{gen} mass distribution originates from the applied cuts. The predictions of the neural network show deviations at the limits of the mass range. Because there are more events in the higher mass range, this effect can most significantly be seen in the range of 250 GeV to 300 GeV. The deviations at the mass range limits are most probably caused by the distribution of events in the training sample, which does not contain any events below 80 GeV or above 300 GeV. However, the results from the neural network show less of a deviation than that of SVfit. Such a deviation in the mean of the reconstructed di-τ mass would not lead to a wrong result in an analysis as the same function would be applied to both data and simulation.

In Fig. 3, the loss (mean squared error) on the training and on the test sample can be seen. The loss shows no sign of overtraining. Overtraining would lead to a discrepancy between the loss on the training and on the test sample and is caused by the neural network describing minor fluctuations in the training sample instead of the underlying relationship.

In Fig. 4, the relative differences per event between the predictions of the neural network and SVfit are shown for the different decay modes. The relative difference per event is calculated as shown in formula (3).

$$\text{relative difference per event} = \frac{\text{di-}\tau_{gen}\text{ mass } - \text{ di-}\tau\text{ mass}}{\text{di-}\tau_{gen}\text{ mass}}. \quad (3)$$

The mean and the standard deviation of the histograms in Fig. 4 refer to the bias and the resolution of the di-τ mass reconstruction, respectively. The bias is the systematic deviation of the predictions in comparison with the target values, and is independent of the bias used in the neural network, while the resolution is a measure of the accuracy of
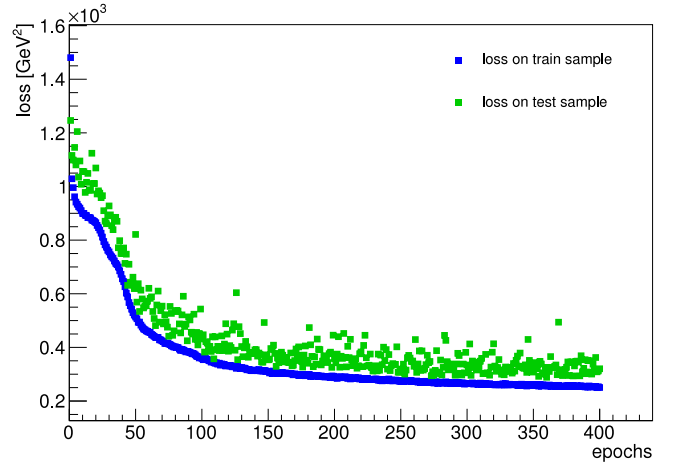
the reconstruction. The bias in the mass determined from the neural network approach using all decay modes is −0.001, which corresponds to only −0.1% of the mean, and is smaller than the bias using SVfit, which is 0.06 or 6% of the mean. The mass resolution determined from the neural network using all decay modes is 0.084 while SVfit has 0.17, which is twice as large. The relative differences are also shown for all the events containing only fully leptonic τ decays, semi-leptonic τ decays, and fully hadronic τ decays. The smallest resolution is achieved for both the neural network and SVfit for events which contain the least number of neutrinos, which are the fully hadronic decays. The resolution is larger for events containing semi-leptonic decays and largest for events which contain fully leptonic decays.

In Fig. 5, the bias of the di-τ reconstruction and its standard error for each di-τ_{gen} mass is shown for the neural network and SVfit in comparison. The neural network and SVfit have their smallest bias around 250 GeV and 100 GeV, respectively.

4.3. Computation time of neural network and SVfit

An important part of the comparison between the neural network and SVfit is the computation time. The benchmark SVfit algorithm requires 6 s of computation time for one event in our computer architecture. Our neural network algorithm requires 110 microseconds per event, a factor of fifty thousand times faster than the SVfit algorithm in the same architecture. The neural network algorithm, however, requires about 17 h to train on 270'000 events with 400 epochs. But once the neural network is trained, the model architecture and the model weights can be saved and used later to make new predictions. In a different computing architecture, using a GPU accompanied by Tensorflow [16] as the backend, the computation time to train the neural network was reduced down to 1.4 h, with a prediction time of 7 μs per event.

4.4. Discrimination between Higgs boson Signal and Drell–Yan background using NN and SVfit

A better discrimination between signal and background events can be achieved by an improved performance of the mass reconstruction, which can be estimated using the metric:

$$\text{Signal to background significance} = \frac{\epsilon_S \sigma_S \mathcal{L}}{\sqrt{\epsilon_B \sigma_B \mathcal{L}}}, \quad (4)$$

to ascertain the signal significance for finding a signal of known cross section σ_S and efficiency ϵ_S over that of a known background, σ_B and ϵ_B, for a given integrated luminosity, $\mathcal{L}$. Fig. 7 shows the mass
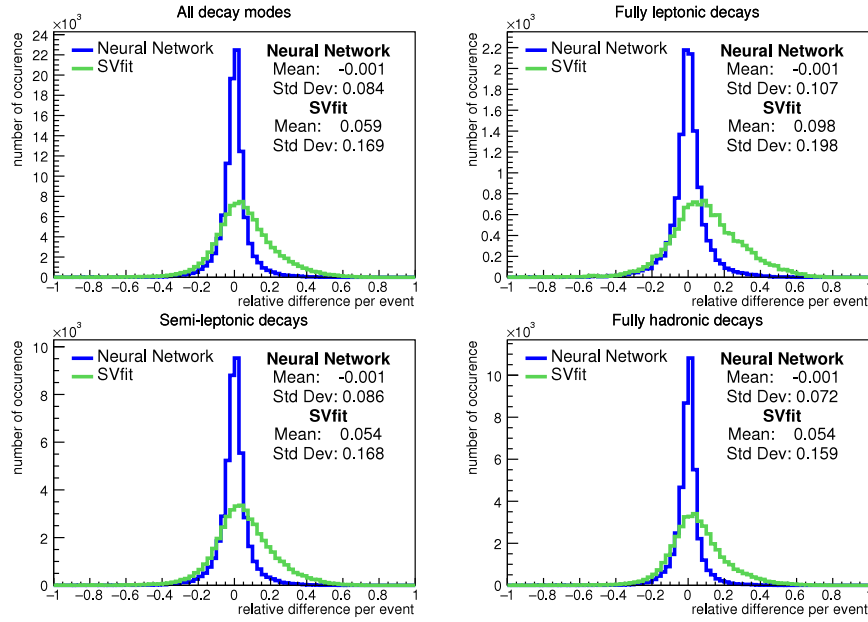
Fig. 4. Relative differences per event, as defined in the text, between the generated mass and the calculated mass determined from both the neural network and SVfit approaches.
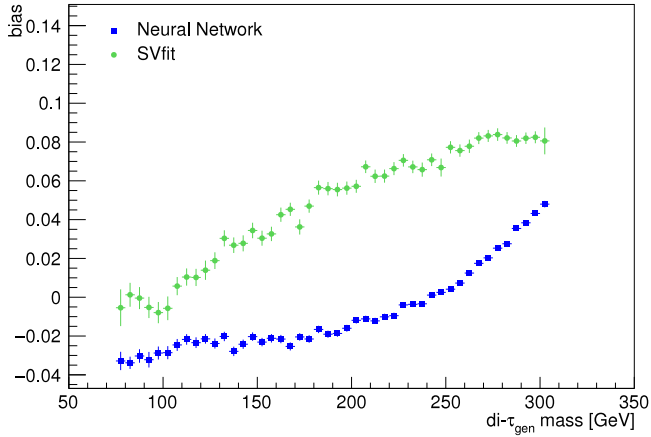

Fig. 5. Bias in the di-τ mass reconstruction per di-τ_{gen} mass.

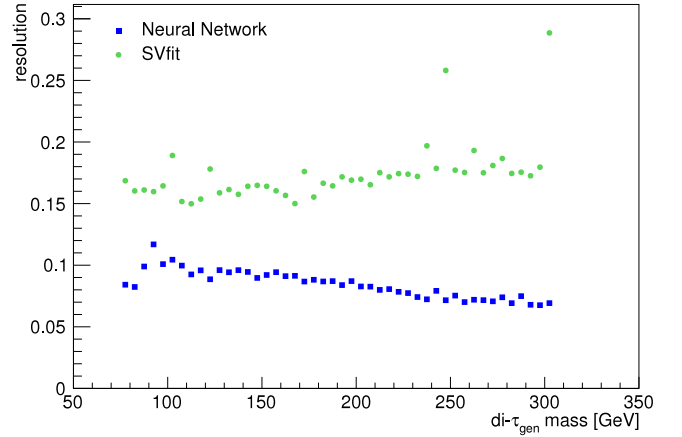**Fig. 6.** The resolution of the di-τ mass reconstruction per di-τ_{gen} mass.

reconstruction for a Higgs boson mass of 125 GeV, consistent with the CMS and ATLAS observations [18,19], also shown with the DY background peaking at the Z boson mass. The Higgs boson signal and the DY background have the same number of events and are not normalized according to their cross section. The signal to background significance in the range of 115–135 GeV for SVfit is 11.2 ± 0.1 and for the neural network 16.5 ± 0.2, where the cross section taken for the DY background is $\sigma_B = 1418 \pm 1$ pb and the cross section assumed for the production of a 125 GeV mass Higgs boson is $\sigma_S = 0.7002 \pm 0.0006$ pb for a center-of-mass energy of 13 TeV, both as obtained with MadGraph [9]. The integrated luminosity is set to a value of $\mathcal{L} = 100$ fb^{-1}, similar to the data sets collected by the ATLAS and CMS experiments until the end of 2017. The efficiency ϵ is the number of events in the signal range divided by the total number of events passing the requirements in Section 2. The DY background events are slightly biased towards higher masses since the neural network is only trained for masses of 80 GeV and above. Using a larger mass range of the training sample would lead to a smaller deviation for the DY background and therefore to an even larger signal to background significance. Fig. 6 shows the comparison of the resolution per di-τ_{gen} mass using either the neural network or SVfit. The resolution of both the neural network and SVfit is constant over the mass range and is larger for SVfit as already shown in Fig. 4.
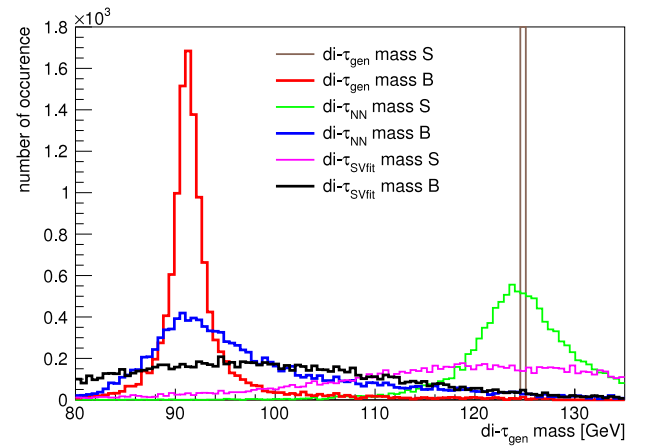

Fig. 7. Reconstructed di-τ mass for a generated Higgs boson (S) and Drell–Yan background (B). Di-τ_{gen} refers to the simulated mass values, di-τ_{NN} and di-τ_{SVfit} to the reconstructed mass values using the neural network and SVfit, respectively.

4.5. Discussion of neural network and SVfit

In Section 4.2, it can be seen that the reconstruction using a neural network shows a significantly smaller bias and better resolution in comparison with SVfit. The computation time, once the neural network is trained, is much shorter than using SVfit, as shown in Section 4.3. Despite those advantages, the neural network has a few drawbacks. The construction of a neural network model, which delivers predictions close to the target values, is time consuming because a slight modification in the model can change the predictions drastically. This is due to the fact that the neural network is trained multiple times with a large sample of events. Unlike SVfit, a large sample of simulated events is required for training the neural network.

Using a neural network for reconstruction has a lot of potential, which warrants further study. A neural network is, in general, not restricted to one target parameter as used here. For example, the whole 4-vector of the di-τ system could be reconstructed if p_T, η, ϕ and the mass are set as target parameters. If the target parameters vary over different scales, like p_T as compared to η, finding a model, and especially an activation function, which delivers predictions close to the target values for all the target parameters, is more difficult than for just one target parameter.

5. Conclusion

We presented a technique for the reconstruction of the mass of a di-τ system using a neural network. The predictions of the neural network are compared to the results using the SVfit likelihood technique, which is currently used by the CMS Collaboration to reconstruct the di-τ mass. For the neural network, the bias in the reconstructed mass is -0.1% and the mass resolution is 8.4%, whereas SVfit has a bias of 6% and a mass resolution of 17%. The signal to background significance of a Higgs boson signal with a mass of 125 GeV and Drell–Yan background using the neural network is 16.5 ± 0.2 and shows an improvement with respect to SVfit, which has a signal to background significance of 11.2 ± 0.1. Using a neural network instead of SVfit improves the performance of the di-τ mass reconstruction significantly. The neural network predicts the di-τ mass approximately fifty thousand times faster than SVfit. With the neural network technique, one must factor a few hours of retraining each time the detector reconstruction or data conditions change in order to ensure that it performs without bias. This extra computing time is not significant, and is commonplace in other machine learning software algorithms in today's experiments. Using a carefully-optimized neural network for reconstruction of the invariant mass of di-τ resonances has significant advantages over common approaches, delivering better mass resolution, reduced bias, and a much faster computation time.

References

[1] CMS Collaboration, Observation of the Higgs boson decay to a pair of τ leptons, Phys. Lett. B 779 (2018) 283, https://arxiv.org/abs/1708.00373.

[2] ATLAS Collaboration, Cross-section measurements of the Higgs boson decaying into a pair of tau-leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, CERN-EP-2018-232, https://arxiv.org/abs/181108856.

[3] CMS Collaboration, The CMS experiment at the CERN LHC, J. Instrum. 3 (2008) S08004.

[4] ATLAS Collaboration, The ATLAS collaboration at the CERN Large Hadron Collider, J. Instrum. 3 (2008) S08003.

[5] L. Bianchini, J. Conway, E.K. Friis, C. Veelken, Reconstruction of the higgs mass in $h \to \tau\tau$ events by dynamical likelihood techniques, J. Phys. Conf. Ser. 513 (2014) 022035.

[6] L. Bianchini, B. Calpas, J. Conway, A. Fowlie, L. Marzola, L. Perinni, C. Veelken, Reconstruction of the Higgs mass in events with Higgs bosons decaying into a pair of τ leptons using matrix element techniques, https://arxiv.org/abs/1603.05910v3.

[7] A. Elagin, P. Murat, A. Pranko, A. Safonov, A new mass reconstruction technique for resonances decaying to $\tau\tau$, Nucl. Instrum. Methods Phys. Res. A 654 (2011) 481–489, https://arxiv.org/abs/1012.4686.

[8] F. Chollet, et al., Keras, 2015, https://keras.io.

[9] J. Alwall, R. Frederix, S. Frixione, et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, J. High Energy Phys. 07 (2014) 079, https://arxiv.org/abs/1405.0301.

[10] T. Sjöstrand, S. Mrenna, P. Skands, PYTHIA 6.4 physics and manual, J. High Energy Phys. 05 (2006) 026, https://arxiv.org/abs/hep-ph/0603175.

[11] T. Sjöstrand, S.Ask, J. Christiansen, R. Corke, N. Desai, P. Ilten, P. Skands, et al., An introduction to PYTHIA 8.2, Comput. Phys. Comm. 191 (2015) 159–177, https://arxiv.org/abs/1410.3012.

[12] DELPHES 3 Collaboration, DELPHES 3: a modular framework for fast simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057, https://arxiv.org/abs/13076346.

[13] J. Baglio, A. Djouadi, Higgs production at the LHC, J. High Energy Phys. 03 (2011) 055, https://arxiv.org/abs/1012.0530.

[14] R.K. Ellis, I. Hinchliffe, M. Soldate, J.J. Van Der Bij, Higgs decay to $\tau^+\tau^-$ a possible signature of intermediate mass Higgs bosons at high energy hadron colliders, Nuclear Phys. B 297 (1988) 221–243.

[15] Theano Development Team, Theano: A Python framework for fast computation of mathematical expressions, 2016, https://arxiv.org/abs/160502688.

[16] M. Abadi, et al., Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, 2018, https://www.tensorflow.org. (Accessed on September 1st).

[17] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, 2014, https://arxiv.org/abs/14126980.

[18] CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B 716 (2012) 30, https://arxiv.org/abs/1207.7235.

[19] ATLAS Collaboration, Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B. 716 (2012) 1–29, https://arxiv.org/abs/1207.7214.