I. RESULTS

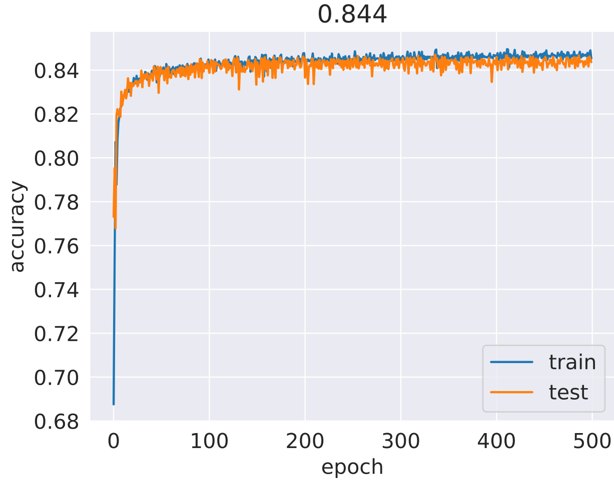
The results that follows are obtained from a grid search where we present best results with the accuracy in a grid with the values of the learning rate $\lambda \in [.....]$ and $\eta \in [1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-5}]$. The other results from the grid search can be found in the GitHub repository[1]. I all of the training of the neural networks for the different datasets explained in Section **??** we use a batch size of 50 000 and train it for 500 epochs.

A. FillMean dataset

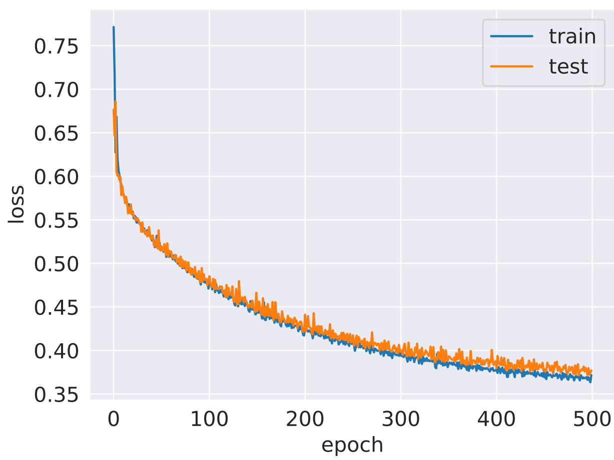
Figure 1 shows the output from the neural network with the dataset *FillMean* described in Section **??** with the accuracy in Figure 1a, loss in Figure 1b and the ROC curve in Figure 1c. The values for the hyperparameters that gave the best results for the neural network from the grid search are $\eta = 1$ and $\lambda = 0.0001$. We see that for both the accuracy and the loss from the training and test data are similar to each other respectively indicating that there is little over training. In addition the accuracy is stable over the epochs while the loss seems to approach a minimum. This indicates that the neural network has learned much of what it can learn from the dataset. The accuracy of the training dataset is 0.844 after the 500th epcoh, while the area under the ROC curve (AUC) in Figure 1c is 0.915.

Figure 2 shows the distribution of the true background and signal events over the output from the neural network in Figure 2a and the log-likelihood ratio of the signal + background and the background estimation $t(\mu)$ in Figure 2a which is shifted by the minimum of $t(\mu)$ with $t(\mu_{min})$. We observe that the distribution of the true background and signal events in Figure 2a are mostly skewed towards 0 and 1 respectively which tells us that most of the events in the dataset has been correctly classified. To calculate the Z score we do a cut at 0.9 as shown by the green dashed line Figure **??**. Counting the background and signal events above 0.9 gives and using Eq. ... gives us $Z = ...$ The estimated likelihood ratio from the neural network is used to find the In Figure 2a we do a cut at 0.9 and obtain $Z_{cut} = 282.852$ when we have used Eq. ... with the number of background and signal events above 0.9 in Figure 2a. From Figure 2c we get the $Z_{likelihood} = 429.963$ by using Eq. ... when $\mu_{min} = 0.607$. The μ value where $t(\mu)$ has it minimum is not at $\mu = 1$ which would be expected from the signal + background hypotosis, and therefore indicates that the neural network has not learned the profile likelihood ratio correctly. The condition for the likelihood ratio estimation from Section **??** was that the loss function should reach a minima in the training, and Figure 1b shows that the loss function seems to reach a minima, but it that it has not converged to one.

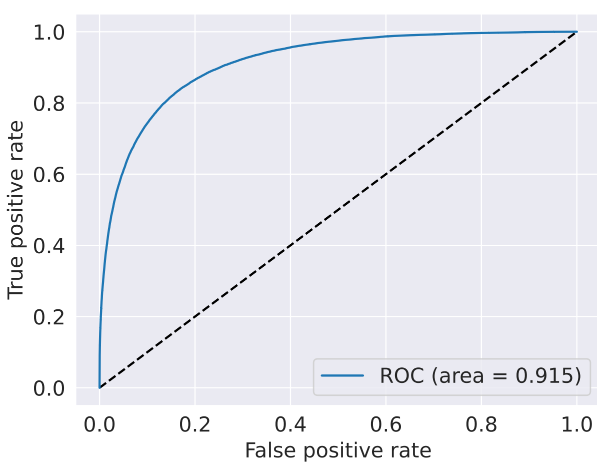
[1] test



(a) Train (blue) and test (orange) accuracy.

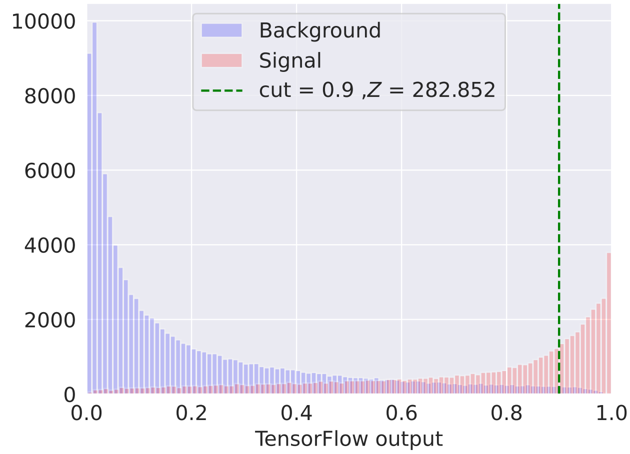


(b) Train (blue) and test (orange) loss.

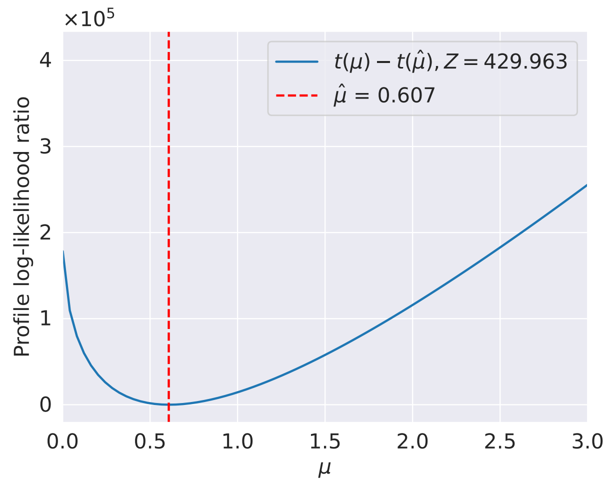


(c) ROC curve of the neural network output.

FIG. 1: The accuracy (a), loss (b) and the ROC curve (c) of the neural network output from the *FillMean* dataset with $\lambda = 1$ and $\eta = 0.0001$.



(a) dd



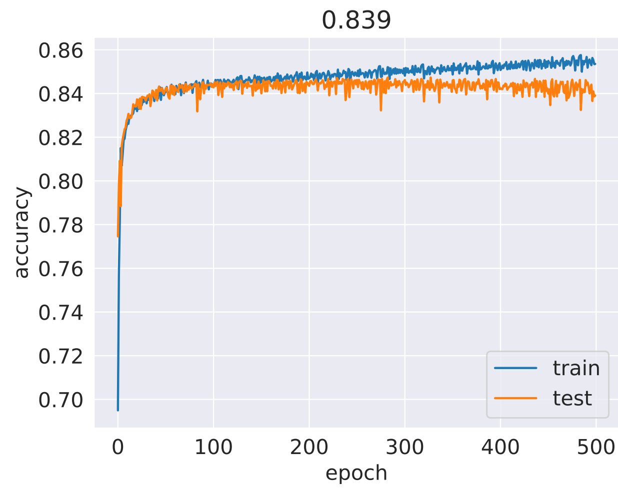
(b) dd

(c)

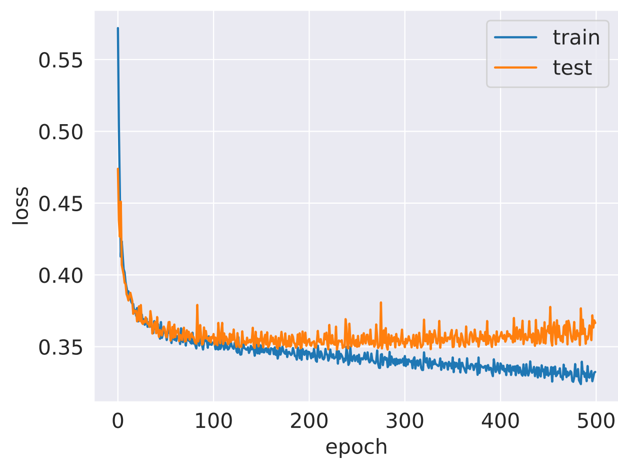
FIG. 2: FillMean

B. FillZero

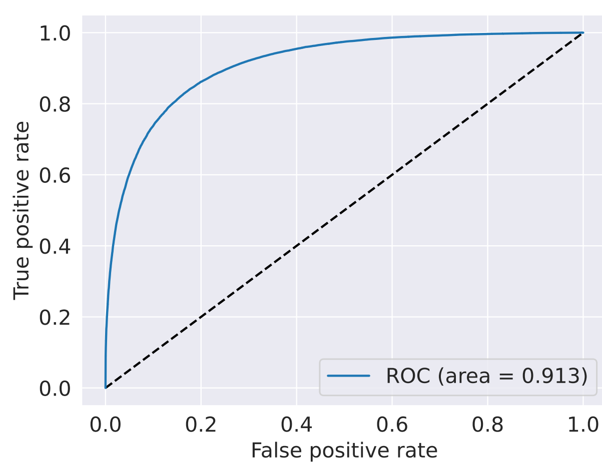
Figure 3 shows the output from the neural network with the dataset *FillMean* described in Section **??** where missing entries are filled with zero and shows the accuracy in Figure 3a, loss in Figure 3b and the ROC curve in Figure 3c. The values for the hyperparameters that gave the best results for the neural network from the grid search are $\eta = 1$ and $\lambda = 10^{-6}$. We see that there is a descrepancy between the values from the training and the testing for both the accuracy and the loss where the accuracy and loss becomes stable for the test dataset, while the accuracy contunies grow and the loss continues to decrease for the training dataset. This can be a sign of the model starting to overfitt the data. The accuracy of the training dataset is 0.839 after the 500th epcoh, while the AUC in Figure 3c is 0.913. Filling the missing entries with zero in particle physics is always not physical since objects, like a jet that is missing would not have any momentum or a direction, and it would therefore cause a bias by inserting a value in the dataset, even if it is zero.



(a) Train (blue) and test (orange) accuracy.



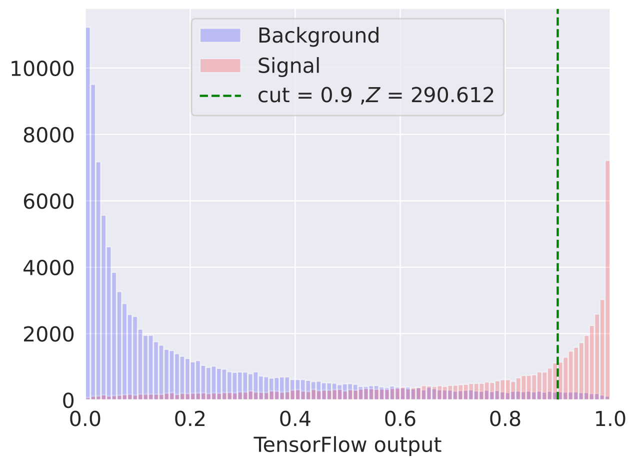
(b) Train (blue) and test (orange) loss.



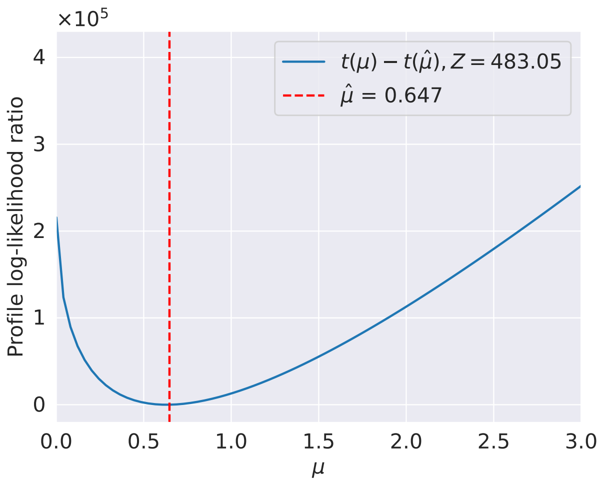
(c) ROC curve of the neural network output.

FIG. 3: The accuracy (a), loss (b) and the ROC curve (c) of the neural network output from the *FillZero* dataset with $\lambda = 1$ and $\eta = 0.0001$.

Figure 4 shows the distribution of the true background and signal events over the output from the neural network in Figure 4a, and the log-likelihood ratio of the signal + background and the background estimation $t(\mu)$ in Figure 4a. We observe that the distribution of the true background and signal events in Figure 4a are mostly skewed towards 0 and 1 respectively which tells us that most of the events in the dataset has been correctly classified. In Figure 4a we do a cut at 0.9 and obtain $Z_{cut} = 290.612$ when we have used Eq. ... with the number of background and signal events above 0.9 in Figure 4a. The log-likelihood function in Figure 4b shifted by the minimum of $t(\mu)$, $t(\mu_{min})$, where the minima of the log-likelihood is located at $\mu_{min} = 0.647$. The minima is not located at $\mu = 1$ which is the expected signal strength under the signal+background hypothosi, and this indicates that the neural network has not learned the likelihood ratio from from the dataset correctly. By using Eq. ..., the $Z_{likelihood}$ is found to be $Z_{likelihood} = 483.050$. The condition for the likelihood ratio estimation from Section **??** was that the loss function should reach a minima in the training, and Figure 3b shows that the loss function seems to reach a minima, but it that it has not converged to one.



(a) dd

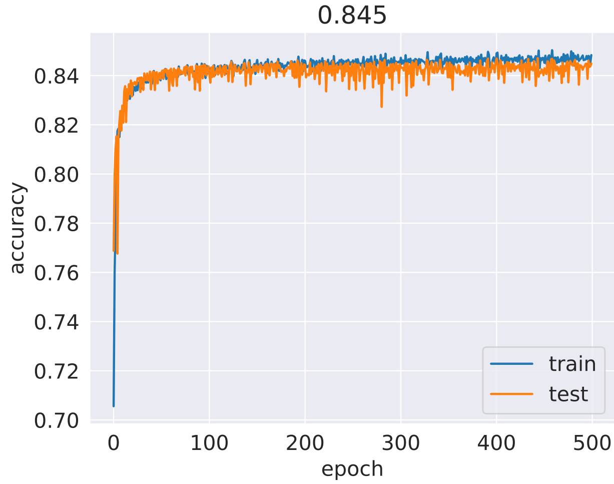


(b) dd

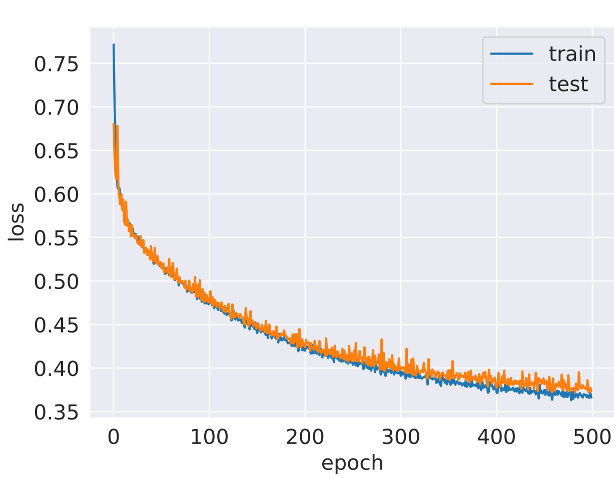
FIG. 4: FillZero

C. FillPhiRandom

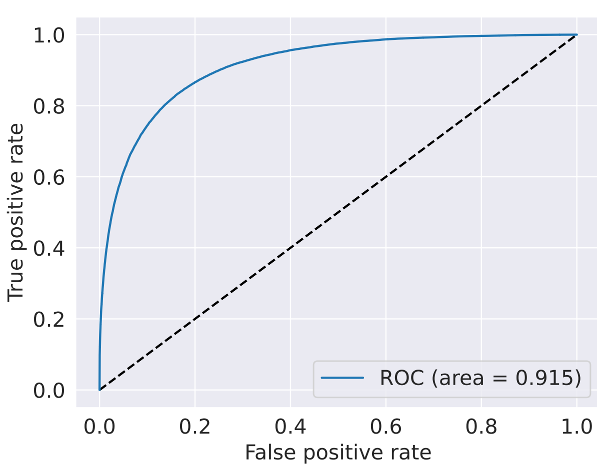
The results from the neural network for the *FillPhi-Random* dataset described in Section **??** are shown in Figure 5 where Figure 5a shows the accuracy, Figure 5b shows the loss, and Figure 3c shows the ROC curve. The values for the hyperparameters that gave the best results for the neural network from the grid search are $\eta = 1$ and $\lambda = 0.0001$. When we compare these results with the ones from the *FillMean* dataset we see that they are almost identical. Both the accuracy and the loss follows the same shape, and the training and testing results are almost the same. The accuracy of the training dataset is 0.845 after the 500th epcoh, while the AUC in Figure 5c is 0.915, which also are almost identical to the results from the *FillMean* dataset. The fact that the results from the training on the *FillMean* and *FillPhiRandom* datasets almost gives the same reslults can be undersood from the fact that the only varialbes that have different values between them is the ϕ variables which are almost uniform as shown in Figure ... in Appendix. This means that filling the missing entries with the mean of that variable or a random number between $-\pi$ and π indicated that the ϕ feautures are not of high importance compared to the other features that the datasets include. From a physics view (if time).



(a) Train (blue) and test (orange) accuracy.

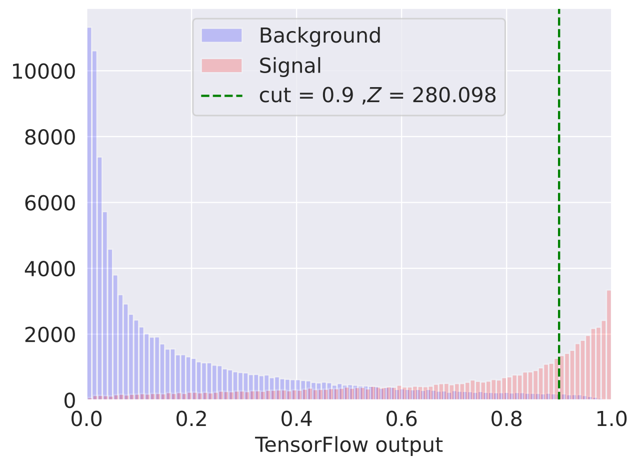


(b) Train (blue) and test (orange) loss.

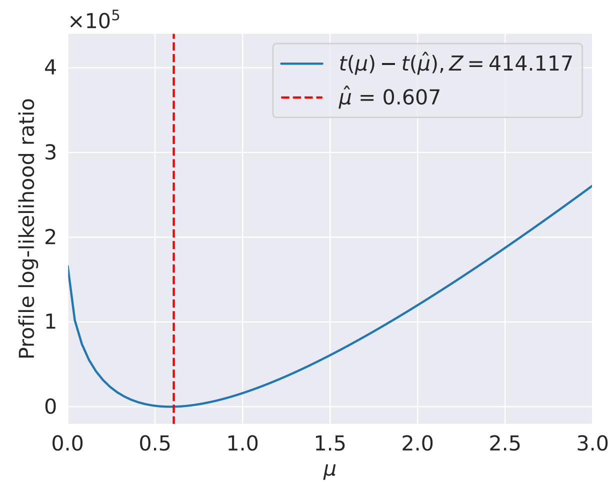


(c) ROC curve of the neural network output.

FIG. 5: The accuracy (a), loss (b) and the ROC curve (c) of the neural network output from the *FillPhiRandom* dataset with $\lambda = 1$ and $\eta = 0.0001$.



(a) dd

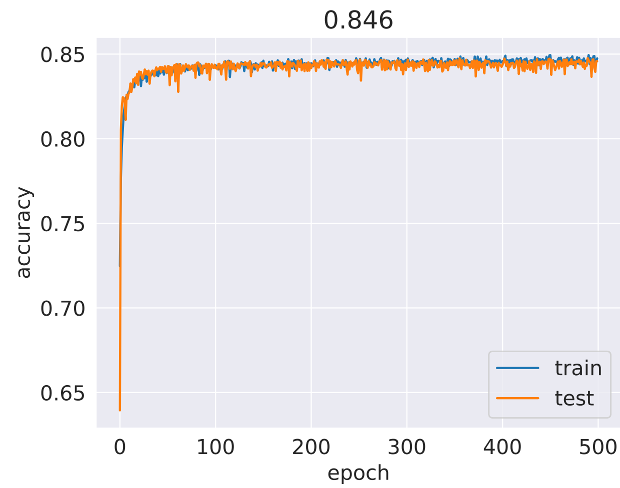


(b) dd

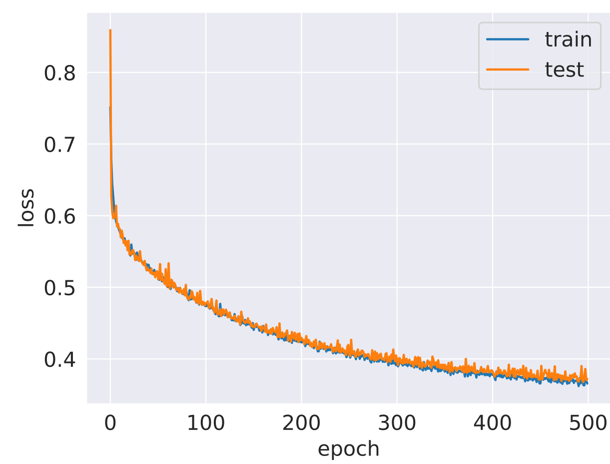
FIG. 6: FillPhiRandom

D. RemovePhi

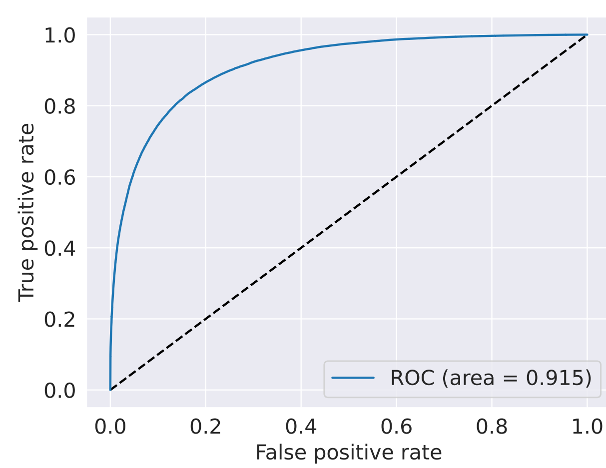
For the dataset *RemovePhi* where we remove the ϕ angle features the results of the neural network are shown in Figure 7a with the accuracy in Figure 7a, loss in Figure 7b and the ROC curve in Figure 9c. The values for the hyperparameters that gave the best results for the neural network from the grid search are $\eta = 1$ and $\lambda = 0.0001$, similar to the training on the *FillMean* and *FillPhirandom* datasets. In this case as well the results from the accuracy, loss and ROC curve are similar to the results for the *FillMean* and *FillPhiRandom* datasets, and accuracy of the training dataset is 0.844 after the 500th epcoh, while the area under the ROC curve (AUC) in Figure 1c is 0.915. This gives us another inducation that the ϕ features are not important for the training since the results are the almost the same when we include them in the dataset. We see that for both the accuracy and the loss from the training and test data are similar to each other respectively indicating that there is little over training. In addition the accuracy is stable over the epochs while the loss seems to approach a minimum. This indicates that the neural network has learned much of what it can learn from the dataset. The accuracy of the training dataset is 0.844 after the 500th epcoh, while the area under the ROC curve (AUC) in Figure 1c is 0.915.



(a) Train (blue) and test (orange) accuracy.

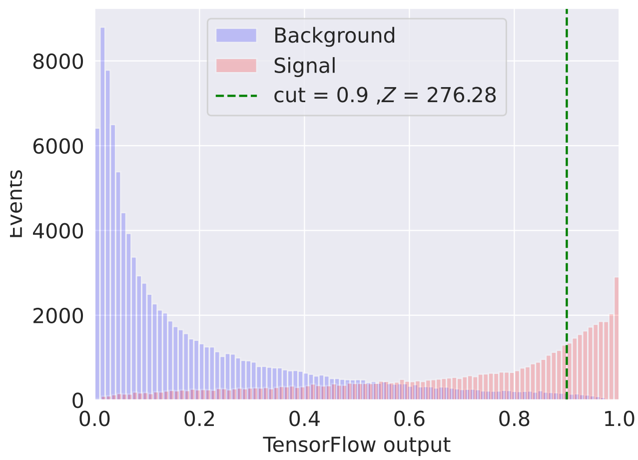


(b) Train (blue) and test (orange) loss.

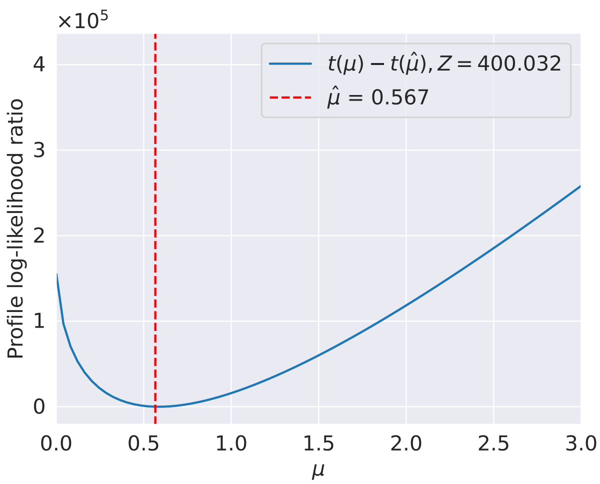


(c) ROC curve of the neural network output.

FIG. 7: The accuracy (a), loss (b) and the ROC curve (c) of the neural network output from the *RemovePhi* dataset with $\lambda = 1$ and $\eta = 0.0001$.



(a) dd

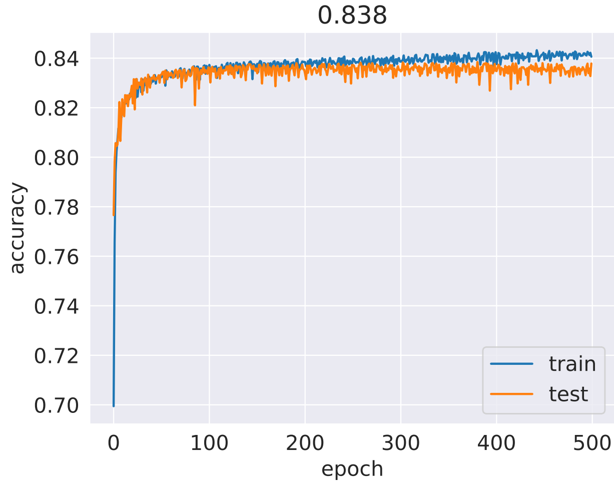


(b) dd

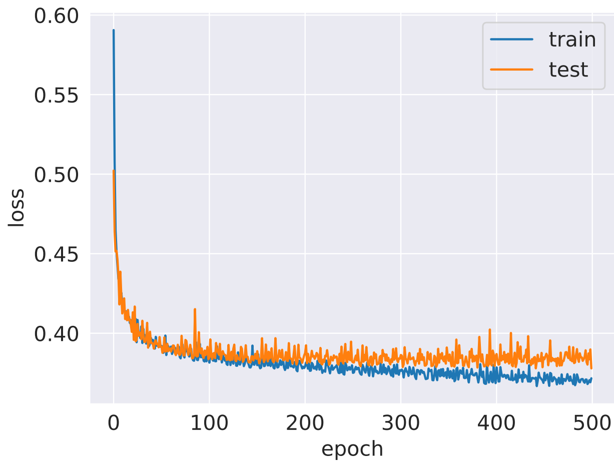
FIG. 8: RemovePhi

E. RemoveJets

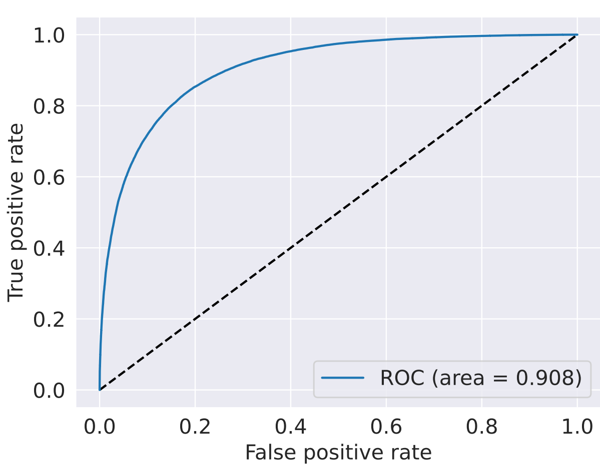
Figure 9 shows the output from the neural network with the dataset *RemoveJets* described in Section **??** where we remove the jets feautures, with the accuracy in Figure 9a, loss in Figure 9b and the ROC curve in Figure 9c. The values for the hyperparameters that gave the best results for the neural network from the grid search are $\eta = 1$ and $\lambda = 0.0001$. We see that for both the accuracy and the loss from the training and test data are similar to each other respectively indicating that there is little over training, but that there are a small descrepancy between them. In addittion to this the accuracy and loss converges to a value which indicates that the model has learned what it can from the dataset. That it converges faster than for e.g *FillMean* comes from the fact that it has fewer input feauters which leads to less information. This shows that in this case it is not an advantage to remove feautures that has missing variables. If the variables had a larger fraction of missing entries then it would maybe advatougus to remove these feautures since any filling of these variables might lead to more noise in the dataset. In addition the accuracy is stable over the epochs while the loss seems to approach a minimum. This indicates that the neural network has learned much of what it can learn from the dataset. The accuracy of the training dataset is 0.844 after the 500th epcoh, while the area under the ROC curve (AUC) in Figure 1c is 0.915.



(a) Train (blue) and test (orange) accuracy.

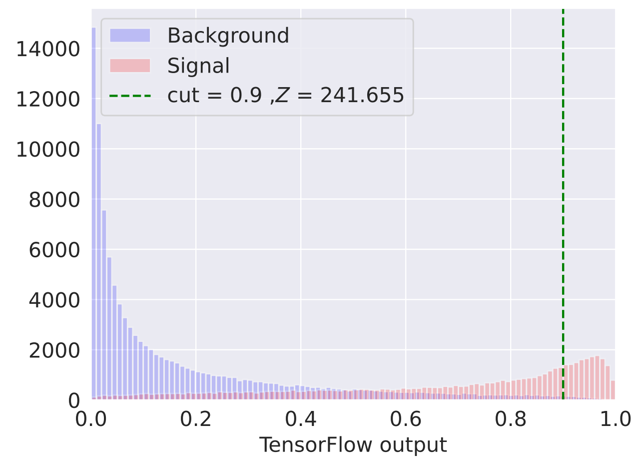


(b) Train (blue) and test (orange) loss.

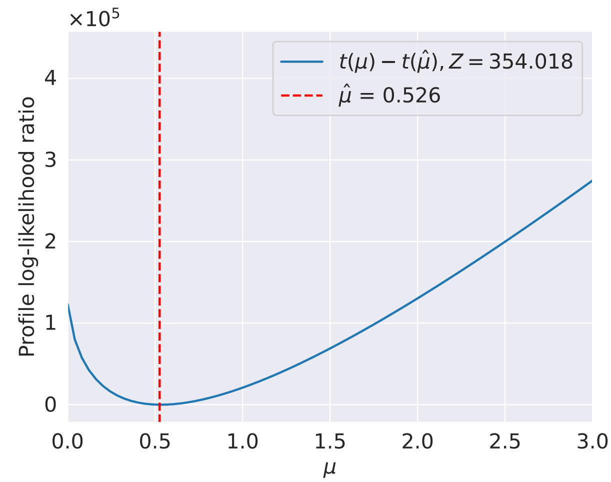


(c) ROC curve of the neural network output.

FIG. 9: The accuracy (a), loss (b) and the ROC curve (c) of the neural network output from the *RemoveJets* dataset with $\lambda = 1$ and $\eta = 0.0001$.



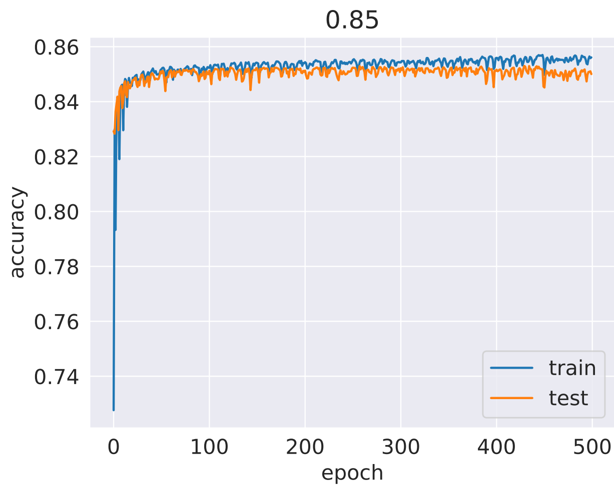
(a) dd



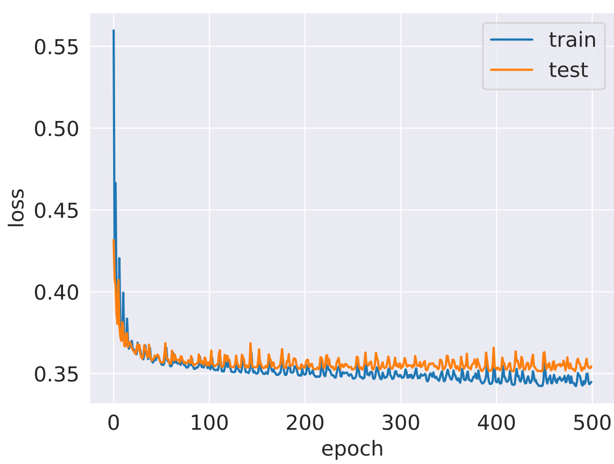
(b) dd

FIG. 10: RemoveJets

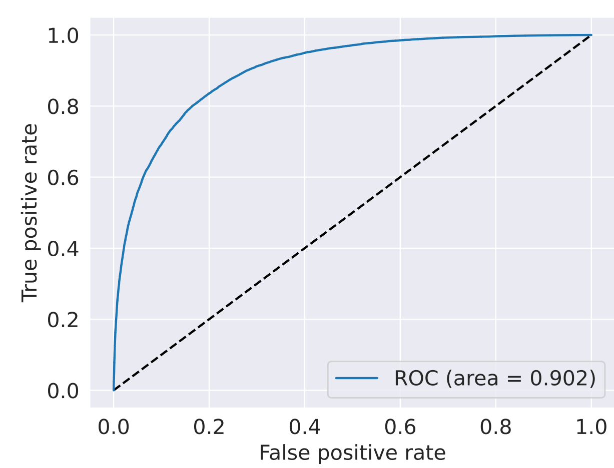
F. JetsNone



(a) Train (blue) and test (orange) accuracy.

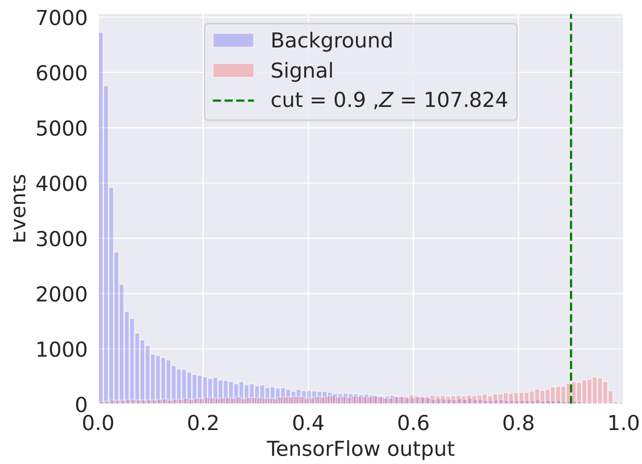


(b) Train (blue) and test (orange) loss.

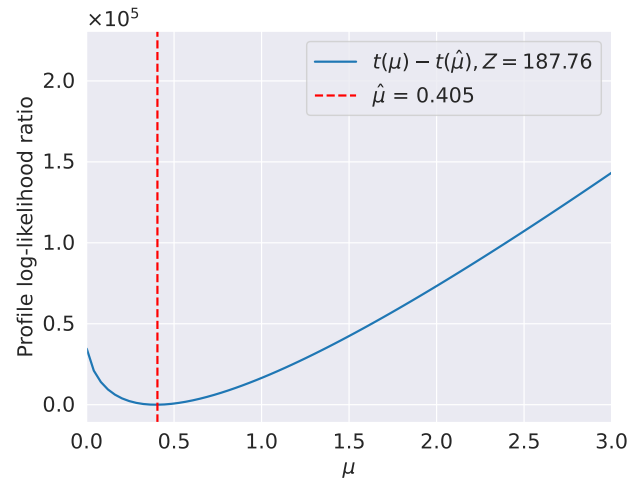


(c) ROC curve of the neural network output.

FIG. 11: The accuracy (a), loss (b) and the ROC curve (c) of the neural network output from the *JetsNone* dataset with $\lambda = 1$ and $\eta = 0.0001$.



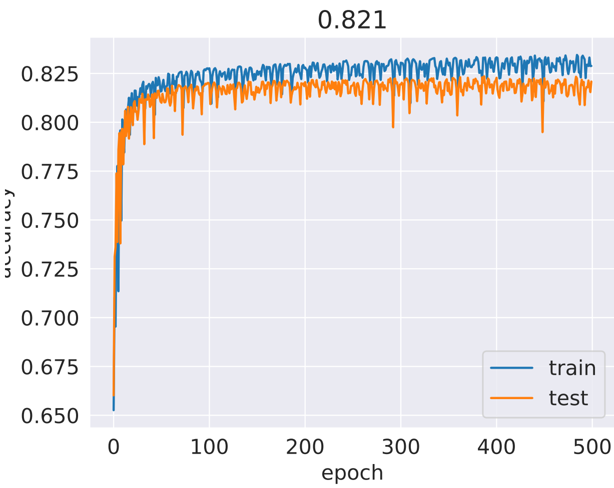
(a) dd



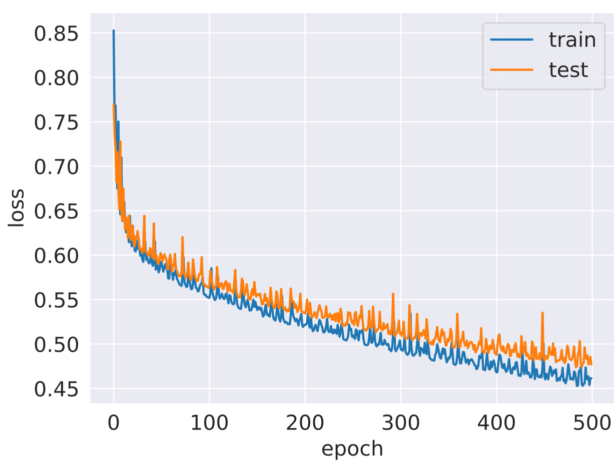
(b) dd

FIG. 12: JetsNone

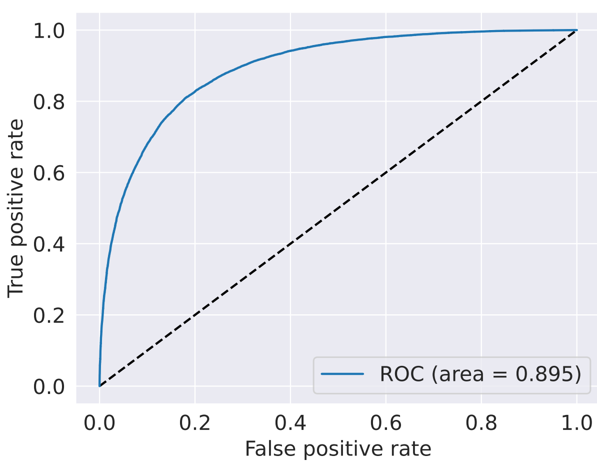
G. JetsOne



(a) Train (blue) and test (orange) accuracy.

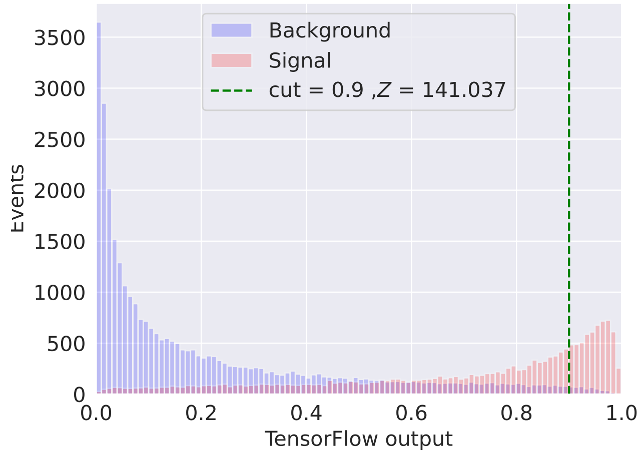


(b) Train (blue) and test (orange) loss.

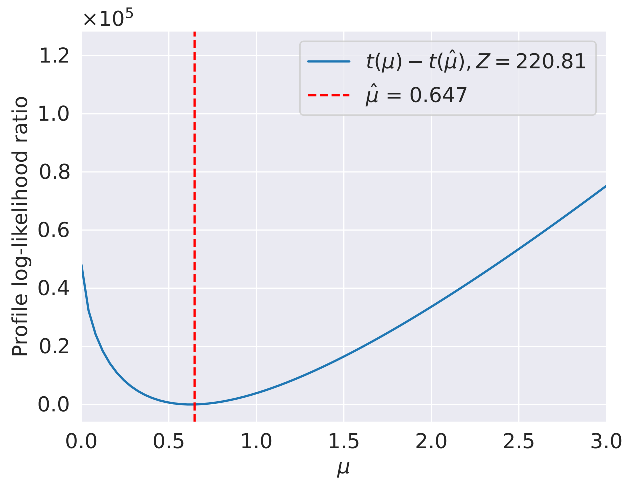


(c) ROC curve of the neural network output.

FIG. 13: The accuracy (a), loss (b) and the ROC curve (c) of the neural network output from the *JetsOne* dataset with $\lambda = 1$ and $\eta = 0.0001$.



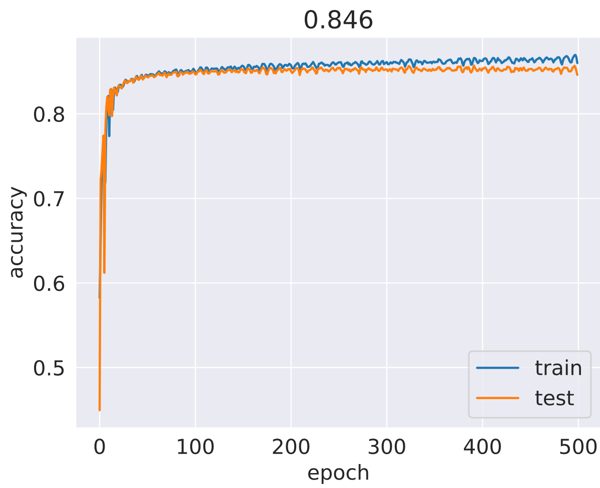
(a) dd



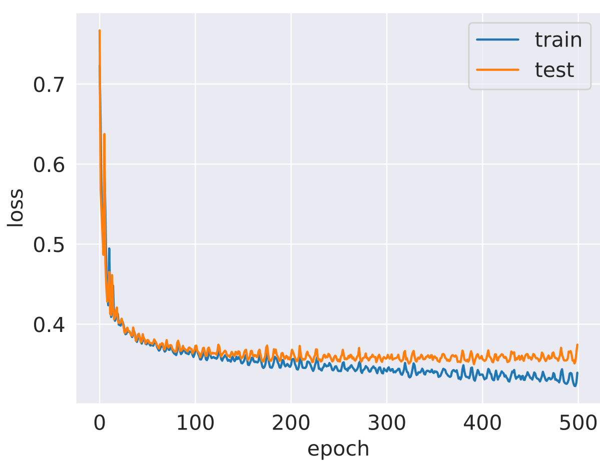
(b) dd

FIG. 14: JetsOne

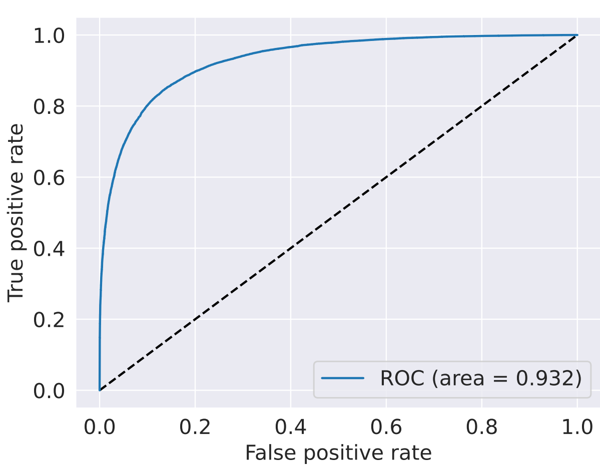
H. JetsTwo



(a) Train (blue) and test (orange) accuracy.

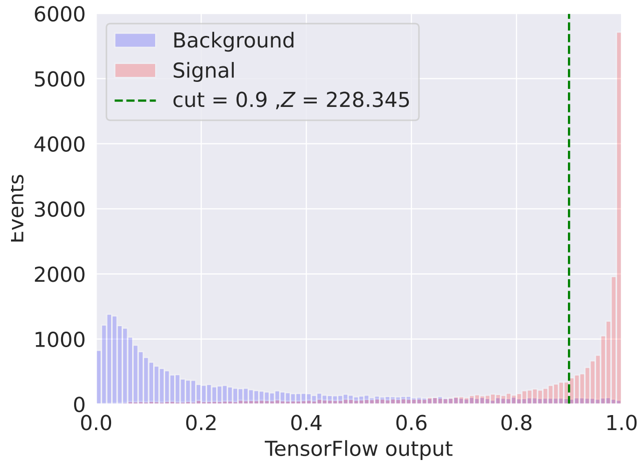


(b) Train (blue) and test (orange) loss.

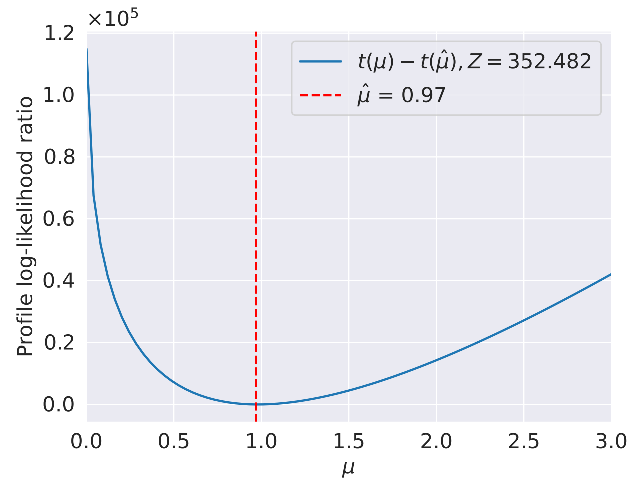


(c) ROC curve of the neural network output.

FIG. 15: The accuracy (a), loss (b) and the ROC curve (c) of the neural network output from the *JetsTwo* dataset with $\lambda = 1$ and $\eta = 0.0001$.



(a) dd



(b) dd

FIG. 16: JetsTwo

To compare and summarize the results we see that among the datasets *FillMean*, *FillZero* and *FillPhiRandom*, both *FillMean* and *FillPhiRandom* gives similar results when we look at the accuracy, loss and the AUC. In the case of *FillZero* it distingushes from the others that it learns the dataset faster and gives a lower AUC compared to the *FillMean* and the *FillPhiRandom* datasets.

When we compare the results from the two datasets *RemovePhi* and *FillMean* we observe that removing the ϕ feautures does not influence the results from the training compared with *FillMean* where we keep the ϕ feautures. This tells us that the ϕ feautures are not among the most important feautures of the Higgs dataset.

In the case of the dataset *RemoveJets* when we have removed the feautures that has missing entries (except MMHiggs) we get a lower performance of the trained model compared to *FillMean* which indicates that the neural network learns from the jet feautures in the dataset and is not influenced to much by the noise when inserting the mean in the dataset. The best result with the accuracy was for the *Jets2* dataset which had the most features among *Jets0*, *Jets1* and *Jets2*, and was also the dataset that was the most balanced with background and signal events. However it did not have the best discovery significanse since it had less statistics than e.g *FillMean*. But the likelihood estimation gave the profile likelihood function that had $\hat{\mu} = 0.97$ which was closest to the expected signal strength equal to 1 indicating that it learned the likelihood ratio better than for the other datasets.

TABLE I: Summary of the results for the different datasets

Dataset	accuracy	AUC	Z_{cut}	$Z_{likelihood}$	μ_{min}
FillMean	0.844	0.915	282.852	429.963	0.607
FillZero	0.839	0.913	290.612	483.050	0.647
FillPhiRandom	0.838	0.908	282.852	429.963	0.607
RemoveJets	0	0.908	282.852	429.963	0.526
RemovePhi	0	0.915	282.852	429.963	0.567
JetsNone	0	0.902	282.852	429.963	0.405
JetsOne	0	0.895	282.852	429.963	0.647
JetsTwo	0	0.932	282.852	429.963	0.970

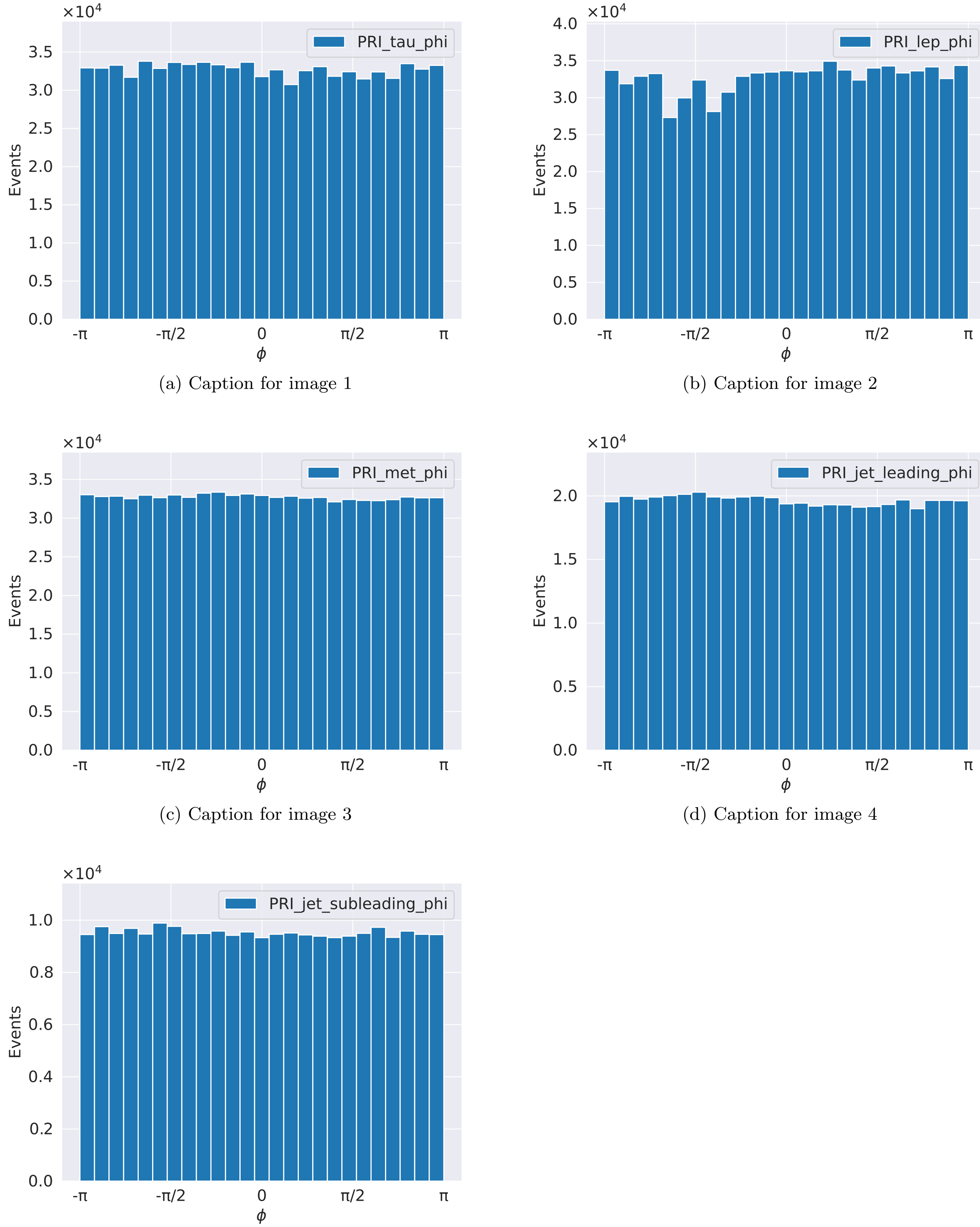
II. CONCLUSION

We have in this project tested various ways of handling the HiggsML dataset which are used to train a neural network, and the output is used to calculate the discovery significanse by using a search region and via likelihood estimation. The dataset that gave the highest discovery significance was the *FillMean* dataset which gave $Z = 429.963$ from the likelihood estimation. However, the *JetsTwo* dataset gave the highest AUC equal to 0.932, but because of less statistics it gave lower discovery significance than *FillMean*. In addition, for all the other datasets except *JetsTwo* the estimated signal strength $\hat{\mu}$ is not close to the expected signal strength equal to 1 indicating that it has not learned the likelihood ratio properly.

III. APPENDIX

TABLE II: Overview of the features in the HiggsML dataset and which of the features that has values in them for the different cases of number of jets.

Features \ n_{jets}	0	1	2/3
EventID	✓	✓	✓
Weight	✓	✓	✓
KaggleSet	✓	✓	✓
KaggleWeight	✓	✓	✓
Label	✓	✓	✓
DER_mass_MMC	✓	✓	✓
DER_mass_transverse_met_lep	✓	✓	✓
DER_mass_vis	✓	✓	✓
DER_pt_h	✓	✓	✓
DER_deltar_tau_lep	✓	✓	✓
DER_pt_tot	✓	✓	✓
DER_sum_pt	✓	✓	✓
DER_pt_ratio_lep_tau	✓	✓	✓
DER_met_phi_centrality	✓	✓	✓
PRI_tau_pt	✓	✓	✓
PRI_tau_eta	✓	✓	✓
PRI_tau_phi	✓	✓	✓
PRI_lep_pt	✓	✓	✓
PRI_lep_eta	✓	✓	✓
PRI_lep_phi	✓	✓	✓
PRI_met	✓	✓	✓
PRI_met_phi	✓	✓	✓
PRI_met_sumet	✓	✓	✓
PRI_jet_num	✓	✓	✓
PRI_jet_all_pt	✓	✓	✓
PRI_jet_leading_pt		✓	✓
PRI_jet_leading_eta		✓	✓
PRI_jet_leading_phi		✓	✓
DER_deltaeta_jet_jet			✓
DER_mass_jet_jet			✓
DER_prodeta_jet_jet			✓
DER_lep_eta_centrality			✓
PRI_jet_subleading_pt			✓
PRI_jet_subleading_eta			✓
PRI_jet_subleading_phi			✓



(a) Caption for image 1

(b) Caption for image 2

(c) Caption for image 3

(d) Caption for image 4

(e) Caption for image 4

FIG. 17: Main caption for the figure.