VBF and ggF Event Discrimination: Using a particle flow neural network to tag VBF events in the $HH \to bbbb$ boosted ATLAS analysis

Gabriel Macdonald*
Supervisor: Max Swiatlowski†

(ATLAS Collaboration)

(Dated: August 21, 2026)

Abstract

In the $HH \to bbbb$ ATLAS analysis, there are two production channels that events must be sorted by: vector boson fusion (VBF) and gluon gluon fusion (ggF). This project utilizes a deep set neural network, the particle flow network, to improve this separation. Two models are built and presented. The first, SRJ PFN, utilizes the small radius jet (SRJ) kinematics as information to discriminate the events. This model increases the VBF efficiency by 12.3 percentage points at the same current ggF HH mis-tag rate, using physics simulations of SM parameters except $\kappa_{2V} = 0.5$. It also improves over a baseline recurrent neural network (RNN), by beating its AUC score by 0.026. A second model, the Higgs PFN, uses the same information but with the addition of two Higgs jet kinematics. This model increases the VBF efficiency by 22.5 percentage points at the same ggF HH mis-tag rate and beats the RNN AUC by 0.09, using the same physics simulations. The models' impact upon the overall analysis and statistical limits is yet to be shown.

* gjmacdon@ualberta.ca; University of Alberta, Canada
† TRIUMF, Canada

I. INTRODUCTION

The Standard Model of particle physics (SM) establishes the theoretical basis for our current understanding of fundamental particle physics [1–3]. It has successfully predicted many different observations in experiments, including the observation of the Higgs Boson in 2012 [4, 5]. The prediction for the Higgs Boson comes from the Brout-Englert-Higgs mechanism, which proposes a spontaneous electroweak symmetry breaking of the Higgs Potential [6–8]. The result of this symmetry breaking leads to a scalar field, whose quantum excitations are the Higgs boson.

Many different parameters and properties have been measured and agree with the SM; however, there are still many things that are yet to be confirmed by experiment. For example, the SM predicts the Higgs boson trilinear self-coupling (HHH) and a coupling between two vector bosons and two Higgs bosons (VVHH, with V = W,Z), but their strengths have not been precisely measured. These couplings are important in the production of Higgs boson pairs, or diHiggs events, because the cross section depends quadratically on both [9]. These couplings can be characterized relative to their SM prediction with values of κ_λ and κ_{2V} for HHH and HHVV, respectively. Precise measurements of κ_λ can inform us of the global structure of the Higgs potential, which can have profound consequences in cosmology, dark energy, dark matter, and matter-antimatter asymmetry [10].

The two main production channels of diHiggs events are vector boson fusion (VBF) and gluon-gluon fusion (ggF). Both of the cross sections for these channels are very low [9], and as such, diHiggs analysis is statistically limited by the number of signal events; because of this, gathering and retaining signal events is of high priority. One of the best final states for analysis is the $HH \to bbbb$, as it has the highest event rate of diHiggs events [11]. In this final state, two analysis methods exist, boosted and resolved. The boosted analysis involves Higgs bosons with high transverse momentum (p_T). We reconstruct each Higgs boson in one large radius jet (LRJ) using a machine learning-based double b tagging process [12, 13]. The resolved analysis is different as each Higgs boson is reconstructed as two separate b tagged small radius jets (SRJ).

The VBF channel cross section is highly sensitive to κ_{2V}, due to the destructive interference between the Feynman diagrams seen in Fig. 1. Changing κ_{2V} disrupts the cancellation between the two different VBF HH amplitudes. For example, in the non SM case where

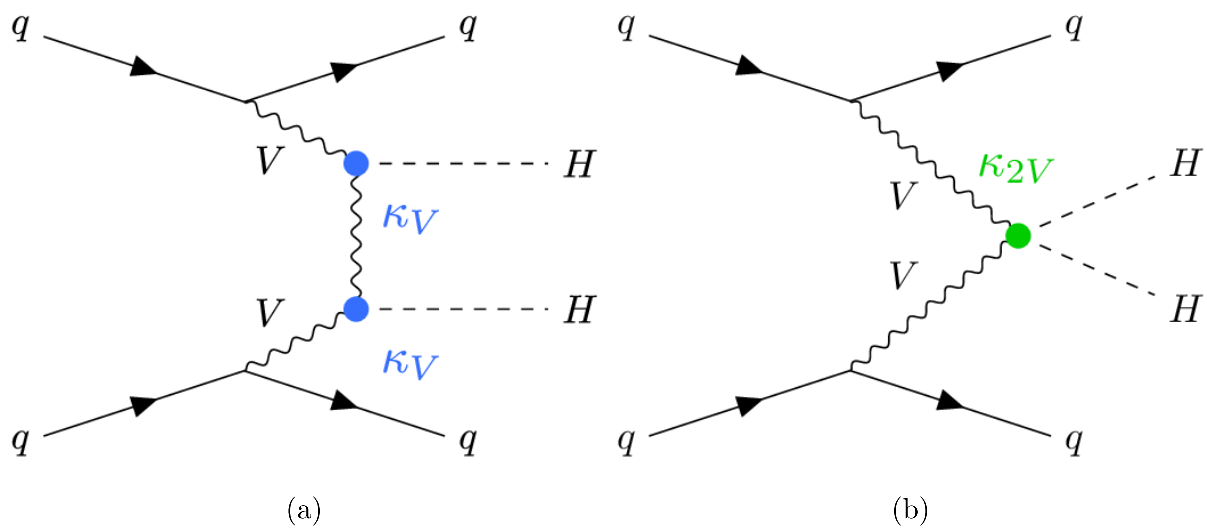


FIG. 1: Two of the leading order Feynman diagrams for VBF Higgs boson pair production. Diagram (a) only involves the Higgs coupling to vector bosons, κ_V, while diagram (b) involves the two Higgs two vector boson coupling, κ_{2V}. These two diagrams interfere destructively. Figure taken from [11].

$\kappa_{2V} = 0$, the cross section for VBF is over 15 times that of the SM [11]. Another significant difference in non SM scenarios is that we expect the Higgs bosons produced to be of higher energy and have a high p_T, making them more central in the detector [14]. This makes the boosted analysis an effective tool for these searches.

Currently in the boosted analysis, to identify the VBF production channel diHiggs events, we look for the presence of two SRJs which do not overlap with the two LRJs. VBF events always produce two SRJs in addition to the Higgs jets due to outgoing quarks produced in the VBF process. We call these non-overlapping SRJs the VBF candidate jets. There are an additional two cuts: a minimum combined mass for the two VBF candidate jets of $m_{jj} > 400$ GeV and a minimum separation of $\Delta\eta > 2$. If these three selections are not passed, events are classified as ggF. The current efficiency of this separation can be seen in Table I. Newly proposed cuts of $m_{jj} > 800$ GeV and $\Delta\eta > 3$, which should improve background rejection in the analysis, can be seen in Table II. Both of these separation techniques lose a substantial amount of signal; for example the VBF channel in the current analysis loses about one third of the total possible signal. This gives us good motivation to find a better technique for tagging the production channel of each event.

The major reason events are falsely tagged is due to SRJs. These jets may fail the detector

TABLE I: Confusion matrix of the VBF and ggF tagging rates using the default SRJ cuts of $m_{jj} > 400\text{GeV}$ and $\Delta\eta > 2$ and Monte Carlo data for all runs from 2015 to 2024 (all SM except $\kappa_{2V} = 0.5$).

		Tagged as	
		VBF	ggF
Actual	VBF	65.6%	34.4%
	ggF	12.2%	87.8%

TABLE II: Confusion matrix of the VBF and ggF tagging rates using the proposed SRJ cuts of $m_{jj} > 800\text{GeV}$ and $\Delta\eta > 3$ and Monte Carlo data for all runs from 2015 to 2024 (all SM except $\kappa_{2V} = 0.5$).

		Tagged as	
		VBF	ggF
Actual	VBF	56.9%	43.1%
	ggF	5.4%	94.6%

acceptance, reconstruction, or selection. This results in VBF events being tagged as ggF due to missing SRJs. Another case is when there are additional, non-VBF SRJs which can come from ISR/FSR and other hard-scatter radiation, as well as from pile-up. This can result in a ggF event being tagged as VBF.

This project aims to improve the separation efficiency of events in the diHiggs boosted channel by using more information from the kinematics of jets. The hope is this will reduce the errors described, and lead to better analysis results. A recurrent neural network (RNN) has been developed for general VBF or VBS (vector boson scattering) event identification in any analysis [15], so we also aim to beat the performance of this model.

II. METHODS

The current analysis and separation technique uses little information about the reconstructed jets within the event. We aim to improve the tagging process of these events by using more of the information coming from the reconstructed kinematics of the SRJ and LRJs, going beyond simple cuts. For example, events with two VBF candidate jets, the reconstructed di mass and $\Delta\eta$ of the jets are both on average higher in VBF events than in ggF events; There is likely correlations between the two that a simple cut does not consider. The current VBF ggF separation does not use any information about the LRJs, but it can be very powerful – especially when we cannot extract information out of the VBF candidate jets (ie, if there are none). We can use the kinematics of the LRJs to help identify VBF

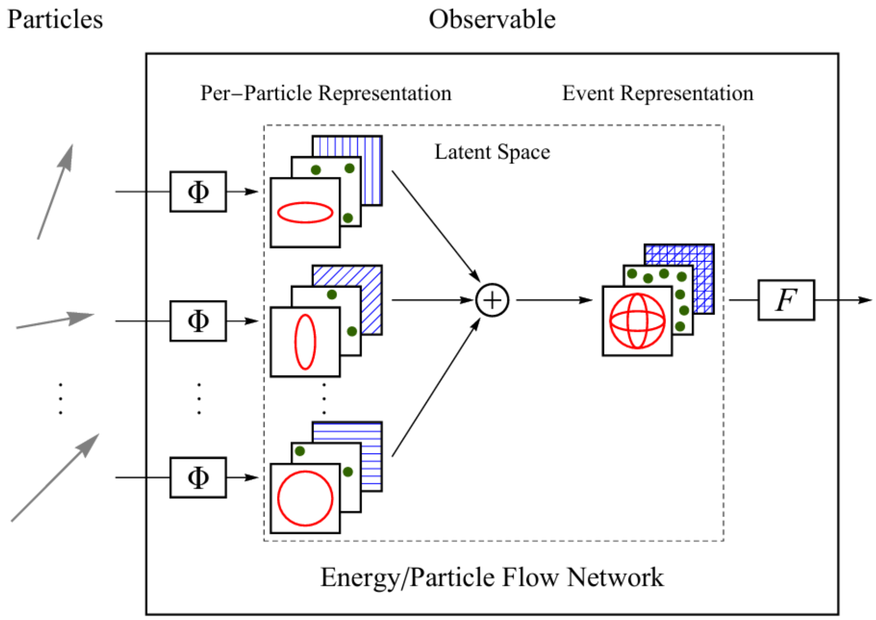


FIG. 2: Visual demonstration of how the PFN proesses particle kinematics, or in our case jet kinematics. They are passed into the ϕ network, which outputs in latent space variables, creating a per-particle or jet representation. These representations are then summed together to build picture of the entire event, which is then processed by the F network and a prediction is returned. Figure taken from [16].

events that may have lost one or both VBF jets down the beamline.

A. Particle Flow Neural Network

To extract the additional information out of the event kinematics, we develop a neural network based on a deep set topology and using the particle flow network (PFN) [16]. This network is formed by two sub networks. The first, which we call the ϕ network, is a simple deep neural network. It takes the kinematic vector of one jet as an input, and outputs many different variables, variables which define a latent space. Each jet is passed through the ϕ network, and then each jet's latent variables are summed together. This allows the overall network to image the global event, and see how each jet is related to each other. For a visualization of this see Fig 2. These summed latent space variables are then passed to another simple deep neural network, called the F network, which passes out the final prediction for whether the event is VBF or ggF.

This network is suitable for our application for two main reasons. Firstly, the pfn will be able to develop and learn patterns in the whole event picture, allowing it to gather information from events with abnormal VBF candidate jet count ($\neq$0 and $\neq$2 for ggF and VBF, respectively). This should improve the tagging of such events. Secondly, the network easily handles variable length input lists; therefore, events with only one SRJ and ones with four or more can be handled the same way within the network. This keeps the network very simple and easy to understand.

We build two models using this architecture. One where we only use the SRJs' kinematics, which we call the SRJ PFN. The other we also use the LRJs' kinematics, in order to gain more information by using the Higgs kinematics, which we call the Higgs PFN.

B. SRJ PFN

We first build a model that only uses the SRJs' kinematic information because the current analysis only uses information from the SRJs. By doing this, we are not introducing very new information into the selection step of the analysis, which should keep the effects on the background of the entire analysis to a minimum. Introducing information from the Higgs jets may cause the selection process to be inconsistent or unpredictable in the background, which can negatively impact the rest of the analysis.

The main intention of this model is for it to learn how to identify events that have additional or extra SRJ not coming from the diHiggs process. This should heavily reduce the number of ggF events being tagged as VBF due to the presence of two or more non VBF SRJs. This may be done by learning to identify which SRJs are not VBF jets. The model should also be able to learn patterns of VBF events where one of the VBF jets are lost to the beam line, where they are undetectable. This should decrease the number of VBF events being tagged as ggF due to only one SRJ being detected.

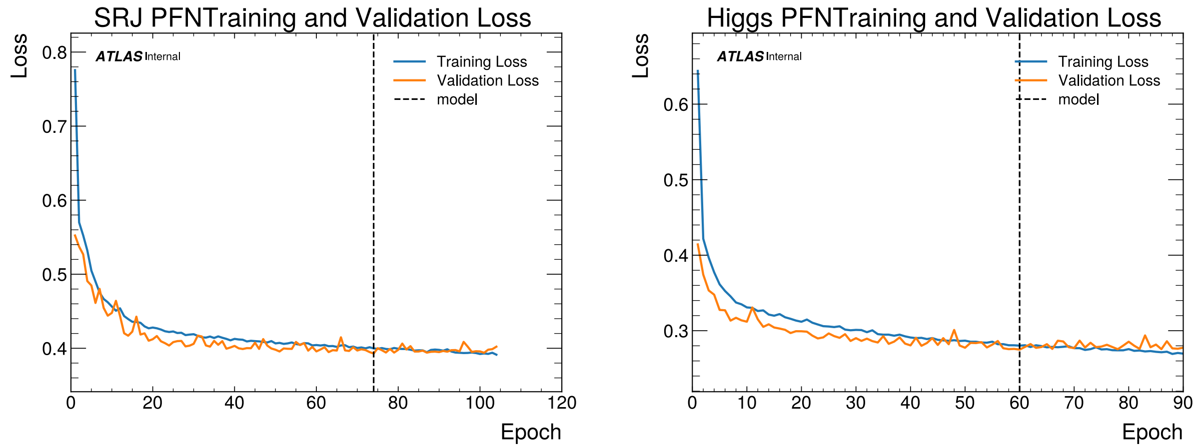
The general structure of the SRJ PFN is as follows. The input takes a list of SRJ three-vectors: (p_T, η, ϕ). It contains all SRJs in the event up to a maximum of four. These jets are sorted in descending p_T in the list. If there are more than four SRJs, than we only take the leading four p_T ones. If there are less than four, then we zero pad the rest of the SRJs spots in the input list. The ϕ network takes the form of three dense layers of 100, 100, and 128, in that order, with a relu activation function. This means there are 128 latent variables. The latent sum is then passed to the F network, which is three dense layers of 100 each, also with a relu activation. The final prediction is outputted using the softmax output activation. The training of the model optimized the cross-entropy loss, using the adam optimizer with a 0.2 dropout in the latent space to prevent overfitting and a 0.001 learning rate. All data was weighted equally at one, except for any jet that was zero padded (no jet), these were weighted at zero. The model was trained until the validation loss did not improve for 30 epochs, at which the parameters of the epoch with the lowest validation loss were restored.

The data chosen for training uses simulated events with $\kappa_{2V} = 0.5$ (SM otherwise), drawn from equally sized random samples of all Monte Carlos from 2015 to 2024. This set of physics parameters was taken as it is near the current lower bound for κ_{2V} [9], and the non SM parameter gives the VBF events a clear signal for the model to learn. We then remove all events with zero SRJs (this slightly changes the cut performance in Tables I and II, and RNN performance, as we are no longer evaluating these events when comparing to the SRJ PFN). The ggF and VBF data is then equalized in count by random selection before training, as to not bias the model to one channel. An 80-10-10 randomized split was then used to create training, validation, and test datasets. The data was trained in batch sizes of 1028, with 42 000 of each ggF and VBF for total training data (not validation or test). The only preprocessing done is to convert all p_T values into TeV units. The final validation loss of the training is 0.4026, and a full training loss curve for the SRJ PFN can be seen in Fig 3a.

C. Higgs PFN

The second model we build uses the same 4 SRJs as the last model, but now also introduces the kinematics of the LRJs. This increases performance further by allowing the model to learn patterns in both the SRJs and LRJs, and even kinematic relations between the two. This even allows the opportunity to tag VBF events, who have zero SRJs, as VBF – if the model learns a distinction in the Higgs kinematics. Doing this is strictly impossible with the current analysis or the last model, as you simply do not have any information from the SRJs.

The inputs for this model are now a list of six four-vectors, with a particle ID (PID) tag as



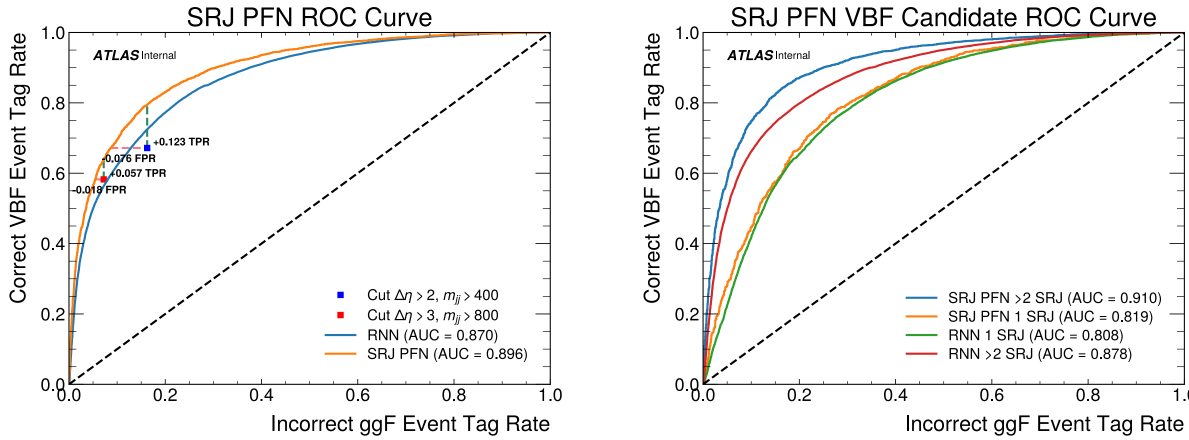
(a) (b)

FIG. 3: Loss plots of the two different models trained. The SRJ PFN loss plot is shown in (a), and the Higgs PFN loss plot is shown in (b). In both of the plots you can see the loss for the training and validation data sets. The "model" black dashed line shows where the model parameters were restored to after 30 epochs of no validation loss improvement.

a fifth entry to each vector. This one hot encode each jet as SRJ or LRJ: $(p_T, \eta, \phi, m, \mathrm{PID})$. The mass (m) of the SRJ's are left as 0. The order of the list is the four SRJs in descending p_T like before, and then the two Higgs jets in descending p_T. The dataset, training parameters, and hyperparameters follow the exact same form as the SRJ PFN, except we now keep zero SRJ events. This increases each VBF and ggF training dataset size to about 57 000. The only preprocessing done is to convert m and p_T to TeV units. The final validation loss of this model is 0.3259, and a full training loss curve for the Higgs PFN can be seen in Fig 3b.

III. RESULTS

The performance of each model can be summarized in a receiver operating characteristic (ROC) curve which works as follows. To use the models we must decide on a network score of which anything scoring higher than that will be classified as VBF, and anything below, ggF. By shifting that cut score from one to zero, we can get any percentage of VBF events labelled as VBF, or the true positive rate (TPR), starting at 0% and going to 100%. As we do this scan, the number of ggF events getting tagged as VBF will also increase (as we loosen the cut), we call this the false positive rate (FPR). This scan can be summarized on



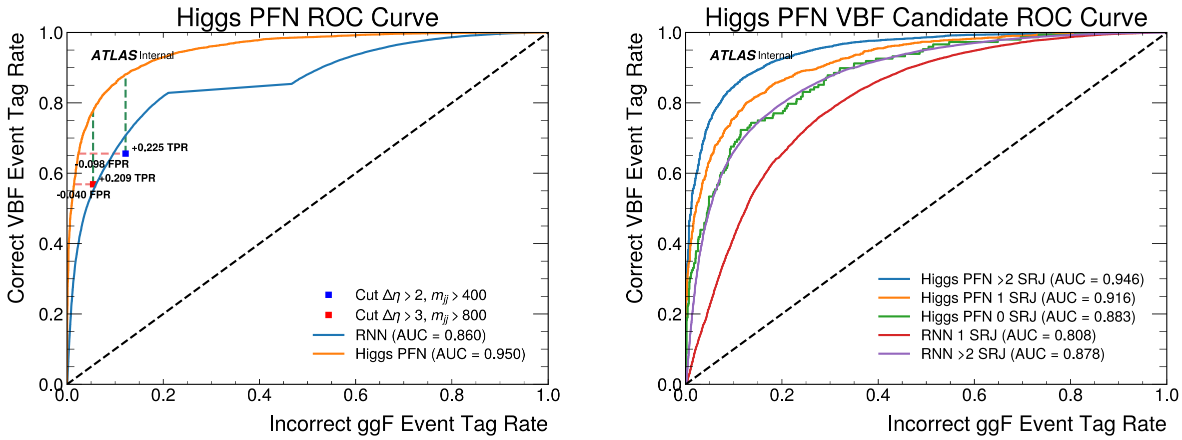
(a) (b)

FIG. 4: ROC curves of the SRJ PFN, using SM physics except $\kappa_{2V} = 0.5$. The overall ROC curve, (a), has an AUC of 0.896, a 0.026 improvement over the RNN. Improvements over the default and proposed cuts for VBF tagging (without 0 SRJ cases, ie different to Tables I and II) are shown on the plot for both the same TPR and decreased FPR, or the same FPR and increased TPR. The ROC curves for events containing 1, and more than 1, SRJ are shown in (b), where we can see the SRJ PFN outperforms the RNN, with great improvement when there is more than 1 SRJ.

a ROC curve, where the best performance is a curve as high and to the left as possible. The area under the curve (AUC) gives a characteristic measure of the model performance, with 1 being perfect and 0.5 being no better than a random selection.

The performance of these models is evaluated using simulations where $\kappa_{2V} = 0.5$ and all other parameters are SM physics. The ROC curves for the SRJ PFN and Higgs PFN can be seen in Fig 4a and Fig 5a, respectively, and can be compared to the RNN ROC curve in each. The SRJ PFN can either improve successful VBF tagging by 12.3 percentage points at the fixed FPR rate of the default analysis cut, or 5.7 percentage points from the proposed cut. Also, it can instead decrease incorrect ggF tagging by 7.6 percentage points at the fixed TPR of the default analysis cut, or 1.9 percentage points from the proposed cut. It beats the RNN AUC score by 0.026. If we separate events into groups which either have one or more than one SRJ, we can make the ROC curves per SRJ case as seen in Fig 4b. This gives us an effective way to see where the performance increase is coming from.

The Higgs PFN shows even more improvement, being able to increase successful VBF



(a) (b)

FIG. 5: ROC curves of the Higgs PFN, using SM physics except $\kappa_{2V} = 0.5$. The overall ROC curve, (a), has an AUC of 0.95, a 0.09 improvement over the RNN. Improvements over the default and proposed cuts for VBF tagging are shown on the plot for both the same TPR and decreased FPR, or the same FPR and increased TPR. The jagged jump in the RNN curve is due to the RNN model assigning the same score for all events with zero SRJs. ROC curves for events which contain 0, 1, or 2 and more SRJs are shown in (b). The model is performing very well with events with any SRJ count, and equalling the RNN 1 SRJ performance for 0 SRJ events.

tagging by 22.5 percentage points at the fixed FPR rate of the default cut, or 20.9 percentage points from the proposed cut, again using $\kappa_{2V} = 0.5$ and all other parameters SM physics. By instead fixing the TPR, it can decrease incorrect ggF tagging by 9.8 percentage points at the default cut and 4.0 percentage points at the proposed cut. Compared to the RNN, its AUC score is 0.09 higher. If we break the events into zero, one, and more than one SRJ cases, we get the ROC curves shown in Fig 5b.

IV. DISCUSSION

By looking at Fig 3 we can see that the features in the training implemented to prevent overtraining, the early stopping and latent dropout, worked. If the training loss dropped significantly below the validation loss, it would indicate overtraining, and memorizing the training dataset. Our loss curves do not do this, and instead the training and validation loss are roughly of the same value, which is indicative of good training.

Overall both of the PFN models are outperforming the two cuts considered in the current analysis, and they beat the RNN. The SRJ PFN can increase VBF tagging by 12.3 percentage points (increasing the efficiency to 79.5%), while maintaining the same rate of ggF events tagged as VBF. This is a substantial improvement over the current 67.2% VBF selection efficiency (ignoring 0 SRJ cases), showing promising results for an improved ggF VBF classifier. This model also outperforms the RNN by improving the AUC by 0.026, a non negligible amount. Because of this, we predict that the SRJ PFN would aid the analysis more than if the RNN was used instead.

The improvement over the RNN comes from better performance when considering events with two or more SRJs. We can see this in Fig 4b, where the SRJ PFN's multi SRJ ROC is outperforming the RNN consistently throughout the curve. This is likely due to the SRJ PFN being able to recognize patterns of background SRJ's being at higher p_T than the VBF jets, as we give the model the top four p_T SRJ's. The RNN, however, only gets the top two.

The Higgs PFN can increase VBF tagging by 22.5 percentage points, while holding the same rate of ggF events tagged as VBF as compared to the default cut. This improves the efficiency to 88.1%. It also beats the RNN by improving the AUC by 0.09. These improvements over the RNN and the SRJ PFN come from increased performance in all SRJ count cases, as seen in Fig 5b. Both cases with SRJs outperform the RNN's two SRJ performance, showing the power of adding the information contained in the Higgs jets. Also showing this power is the performance on zero SRJ cases, where the Higgs PFN is performing equally as good as the RNN two SRJ case. We also see how the RNN is unsure of zero SRJ events, assigning them all the same score, causing a jump in the ROC score in Fig 5a.

The Higgs PFN has larger improvements when compared to the SRJ PFN. Due to this, one would hope to have an even better impact on the analysis and final results over the SRJ PFN; however, this needs to be tested thoroughly, as this model introduces new information previously not used in the selection process. Because of this, it may be the case that using the Higgs jet information causes the multi jet backgrounds to become a nuisance and difficulty in the analysis. For this reason, we are more unsure if the Higgs PFN will have a positive impact on the final analysis.

The two models both show significant and promising results for a better VBF/ggF classifier; however, their impact on multijet background modelling, systematic uncertainties, and expected κ_{2V} limits are yet to be evaluated and determined. Further optimization and improvements can also be made to these two models. None of the hyperparameters of the two PFNs were optimized, so a full sweep and optimization of these could increase our performance. Since the deep set network structure used is also very simple, one can also test further improvements by using larger and more complex models, like a transformer.

V. CONCLUSION

We implemented a deep set neural network to improve VBF and ggF event identification in the boosted $HH \to bbbb$ analysis. By using the particle flow network structure [16], we successfully developed two models, the SRJ PFN and the Higgs PFN, which are trained, tested, and evaluated using SM physics except $\kappa_{2V} = 0.5$. By using more information from the kinematics of the SRJs in the event than the default analysis cut, the SRJ PFN is able to increase successful VBF tagging by 12.3 percentage points, and beat the RNN AUC score by 0.026. This shows that the model is extracting more information out of the SRJ kinematics. By using the same SRJ information, with the addition of the Higgs jet kinematics, the Higgs PFN was able to increase successful VBF tagging by 22.5 percentage points and improve further over the RNN by increasing the AUC by 0.09. This model demonstrates how there is even more information we can extract out of the Higgs jet kinematics to improve event discrimination. While these results show promising classifier level improvements, the impacts of the models on multijet background modelling, systematic uncertainties, and expected κ_{2V} limits remain to be shown.

VI. ACKNOWLEDGEMENTS

I would like to give my greatest thanks and appreciation Dr. Max Swiatlowski and Dr. Dilia María Portillo Quintero for their superb mentorship over the summer and for the future. Their expertise and experience were an amazing tool I am grateful to have had throughout the summer and this project.

I am further grateful to the Dr. Swiatlowski's other students and the members in the $HH \to bbbb$ group for their contributions and comments in meetings. Further thanks goes to TRIUMF and the Institute of Particle Physics in Canada for the support, funding, and travel they provided that made this entire summer possible.

Thank you to TRIUMF and CERN for giving me the opportunity to join in their summer student programs alongside this project. The lectures in both of the programs were excellent and will be a core component of my education.

---

[1] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264

[2] A. Salam, *Weak and electromagnetic interactions*, pp. 244–254. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/9789812795915_0034

[3] S. L. Glashow, "Partial-symmetries of weak interactions," *Nuclear Physics*, vol. 22, no. 4, pp. 579–588, 1961. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0029558261904692

[4] ATLAS Collaboration, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc," *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037026931200857X

[5] CMS Collaboration, "Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc," *Physics Letters B*, vol. 716, no. 1, pp. 30–61, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0370269312008581

[6] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.13.321

[7] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.13.508

[8] P. Higgs, "Broken symmetries, massless particles and gauge fields," *Physics Letters*, vol. 12, no. 2, pp. 132–133, 1964. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0031916364911369

[9] The ATLAS and CMS Collaborations, "Combination of atlas and cms searches for higgs boson pair production at $\sqrt{s} = 13$ tev," 2026. [Online]. Available: https://arxiv.org/abs/2602.23991

[10] S. D. Bass, A. De Roeck, and M. Kado, “The Higgs boson implications and prospects for future discoveries,” *Nature Reviews Physics*, vol. 3, no. 9, pp. 608–624, sep 2021. [Online]. Available: https://doi.org/10.1038/s42254-021-00341-2

[11] ATLAS Collaboration, “Search for pair production of boosted higgs bosons via vector-boson fusion in the bb¯bb¯ final state using pp collisions at s=13tev with the atlas detector,” *Physics Letters B*, vol. 858, p. 139007, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0370269324005653

[12] “Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS,” CERN, Geneva, Tech. Rep., 2020, all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-019. [Online]. Available: https://cds.cern.ch/record/2724739

[13] “Efficiency corrections for a tagger for boosted $H \to b\bar{b}$ decays in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,” CERN, Geneva, Tech. Rep., 2021, all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-035. [Online]. Available: https://cds.cern.ch/record/2777811

[14] F. Bishara, R. Contino, and J. Rojo, “Higgs pair production in vector-boson fusion at the lhc and beyond,” *The European Physical Journal C*, vol. 77, no. 7, p. 481, Jul 2017. [Online]. Available: https://doi.org/10.1140/epjc/s10052-017-5037-9

[15] A. Giannini, V. Del Piano, and C. L. Van Eck, “VBF-RNN: a general tool to solve the VBF/VBS identification task,” CERN, Geneva, Tech. Rep., 2026. [Online]. Available: https://cds.cern.ch/record/2965720

[16] P. T. Komiske, E. M. Metodiev, and J. Thaler, “Energy flow networks: deep sets for particle jets,” *Journal of High Energy Physics*, vol. 2019, no. 1, Jan. 2019. [Online]. Available: http://dx.doi.org/10.1007/JHEP01(2019)121