

Classification of Electron and Muon Neutrino Events for the ESS ν SB Near Water Cherenkov Detector using Graph Neural Networks

J. Aguilar^a M. Anastasopoulos^b D. Barčot^r E. Baussan^c A.K. Bhattacharyya^b
 A. Bignami^b M. Blennow^{d,e} M. Bogomilov^f B. Bolling^b E. Bouquerel^c F. Bramati^g
 A. Branca^g G. Brunetti^g A. Burgman^{h,1,2} I. Bustinduy^a C.J. Carlile^h J. Cederkall^{h,1}
 T. W. Choiⁱ S. Choubey^{d,e} P. Christiansen^{h,1} M. Collins^{j,b} E. Cristaldo Morales^g
 P. Cupia^k D. D'Ago^y H. Danared^b J. P. A. M. de André^c M. Dracos^c I. Efthymiopoulos^l
 T. Ekelöfⁱ M. Eshraqi^b G. Fanourakis^m A. Farrickerⁿ E. Fasoula^{o,p} T. Fukuda^q
 J. García-Marcos^u N. Gazis^b Th. Geralis^m M. Ghosh^r A. Giarnetti^s G. Gokbulut^{t,u}
 C. Hagner^v L. Halić^r M. Hooft^u K. E. Iversen^{h,1} N. Jachowicz^u M. Jenssen^b
 R. Johansson^b I. Karakoulias^{m,o} E. Kasimi^o A. Kayis Topaksu^t B. Kildetoft^b B. Kliček^r
 K. Kordas^{o,p} A. Leisos^w M. Lindroos^{b,h,3} A. Longhin^x C. Maiano^b S. Marangoni^g
 C. Marrelli^l D. Meloni^s M. Mezzetto^y N. Milas^b J.L. Muñoz^a K. Niewczas^u M. Oglakci^t
 T. Ohlsson^{d,e} M. Olvegårdⁱ M. Pari^w J. Park^{h,1,4} D. Patrzalek^b G. Petkov^f Ch. Petridou^{o,p}
 P. Poussot^c A. Psallidas^m F. Pupilli^y D. Saiang^z D. Sampsonidis^{o,p} A. Scanu^g
 C. Schwab^c F. Sordo^a A. Sosa^{aa} G. Stavropoulos^m M. Stipčević^r R. Tarkeshian^b
 F. Terranova^g T. Tolba^y E. Trachanas^b R. Tsenov^f A. Tsirigotis^w S. E. Tzamarias^o
 M. Vanderpoorten^u G. Vankova-Kirilova^f N. Vassilopoulos^{ab} S. Vihonen^{d,e} J. Wurtz^c
 V. Zeter^c O. Zormpa^m

^aConsorcio ESS-bilbao, Parque Científico y Tecnológico de Bizkaia, Laida Bidea, Edificio 207-B, 48160 Derio, Bizkaia, Spain

^bEuropean Spallation Source, Box 176, SE-221 00 Lund, Sweden

^cIPHC, Université de Strasbourg, CNRS/IN2P3, Strasbourg, France

^dDepartment of Physics, School of Engineering Sciences, KTH Royal Institute of Technology, Roslagstullsbacken 21, 106 91 Stockholm, Sweden

^eThe Oskar Klein Centre, AlbaNova University Center, Roslagstullsbacken 21, 106 91 Stockholm, Sweden

^fSofia University St. Kliment Ohridski, Faculty of Physics, 1164 Sofia, Bulgaria

^gUniversity of Milano-Bicocca and INFN Sez. di Milano-Bicocca, 20126 Milano, Italy

^hDepartment of Physics, Lund University, P.O. Box 118, 221 00 Lund, Sweden

ⁱDepartment of Physics and Astronomy, FREIA Division, Uppsala University, P.O. Box 516, 751 20 Uppsala, Sweden

¹Corresponding authors

²Current address: Department of Physics, Stockholm University, 106 91 Stockholm, Sweden

³deceased

⁴Current address: Center for Exotic Nuclear Studies, Institute for Basic Science, 34126 Daejeon, Republic of Korea

^jFaculty of Engineering, Lund University, P.O Box 118, 221 00 Lund, Sweden

^kAGH University of Krakow, al. A. Mickiewicza 30, 30-059 Krakow, Poland

^lCERN, 1211 Geneva 23, Switzerland

^mInstitute of Nuclear and Particle Physics, NCSR Demokritos, Neapoleos 27, 15341 Agia Paraskevi, Greece

ⁿCockcroft Institute (A36), Liverpool University, Warrington WA4 4AD, UK

^oDepartment of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece

^pCenter for Interdisciplinary Research and Innovation (CIRI-AUTH), Thessaloniki, Greece

^qInstitute for Advanced Research, Nagoya University, Nagoya 464–8601, Japan

^rCenter of Excellence for Advanced Materials and Sensing Devices, Ruđer Bošković Institute, 10000 Zagreb, Croatia

^sDipartimento di Matematica e Fisica, Università di Roma Tre, Via della Vasca Navale 84, 00146 Rome, Italy

^tUniversity of Cukurova, Faculty of Science and Letters, Department of Physics, 01330 Adana, Turkey

^uDepartment of Physics and Astronomy, Ghent University, Proeftuinstraat 86, B-9000 Ghent, Belgium

^vInstitute for Experimental Physics, Hamburg University, 22761 Hamburg, Germany

^wPhysics Laboratory, School of Science and Technology, Hellenic Open University, 26335, Patras, Greece

^xDepartment of Physics and Astronomy "G. Galilei", University of Padova and INFN Sezione di Padova, Italy

^yINFN Sez. di Padova, Padova, Italy

^zDepartment of Civil, Environmental and Natural Resources Engineering Luleå University of Technology, SE-971 87 Lulea, Sweden

^{aa}ISIS, STFC, Rutherford Appleton Laboratory, Harwell Oxford, Didcot OX11 0QX, United Kingdom

^{ab}Institute of High Energy Physics (IHEP) Dongguan Campus, Chinese Academy of Sciences (CAS), Guangdong 523803, China

E-mail: alexander.burgman@fysik.su.se, joakim.cederkall@fysik.lu.se,
peter.christiansen@fysik.lu.se, kaare.iversen@fysik.lu.se, jcpark@ibs.re.kr

ABSTRACT: In the effort to obtain a precise measurement of leptonic CP-violation with the ESS ν SB experiment, accurate and fast reconstruction of detector events plays a pivotal role. In this work, we examine the possibility of replacing the currently proposed likelihood-based reconstruction method with an approach based on Graph Neural Networks (GNNs). As the likelihood-based reconstruction method is reasonably accurate but computationally expensive, one of the benefits of a Machine Learning (ML) based method is enabling fast event reconstruction in the detector development phase, allowing for easier investigation of the effects of changes to the detector design. Focusing on classification of flavour and interaction type in muon and electron events and muon- and electron neutrino interaction events, we demonstrate that the GNN reconstructs events with greater accuracy than the likelihood method for events with greater complexity, and with increased speed for all events. Additionally, we investigate the key factors impacting reconstruction performance, and demonstrate how separation of events by pion production using another GNN classifier can benefit flavour classification.

KEYWORDS: Neutrino Detectors, Cherenkov detectors, Analysis and statistical methods, Performance of High Energy Physics Detectors

Contents

1	Introduction	1
2	Methods	3
2.1	Simulation	3
2.2	Reconstruction	3
2.2.1	Graph Neural Networks	3
2.2.2	Likelihood-based Reconstruction	4
2.2.3	Training	4
2.2.4	Performance Metric	4
2.3	Likelihood-based Cuts	6
3	Results	7
3.1	Charged Lepton Events	7
3.2	Charged Current Neutrino Events	8
3.2.1	With Likelihood-based Cuts	8
3.2.2	Without Likelihood-based Cuts	10
3.2.3	Performance Comparison	11
3.2.4	Pion Production	14
3.2.5	Pion Classifier	14
3.2.6	Flavour Classification Performance Split by the Pion Model	15
3.3	Neutral Current Neutrino Events	15
3.3.1	Separation of CC and NC events	17
3.4	Reconstruction Speed	18
4	Conclusion	18
A	Events Displays of Events with Superior GNN Performance	19

1 Introduction

The existence of CP violation in the leptonic sector is among the most interesting open questions in particle physics, and can be investigated through the study of neutrino oscillations [1]. The potential impact of the CP violating phase δ_{CP} on the neutrino oscillation probability is almost 3 times larger at the second oscillation maximum than at the first [2], but requires 9 times more statistics, as the second maximum is located 3 times further away than the first, and the flux is inversely proportional to the square of the distance. This can be achieved by using a more intense neutrino beam (about one order of magnitude), measuring over a longer period (about 9 times longer), or using more efficient event selection algorithms. The European Spallation Source (ESS) linear accelerator (linac) in

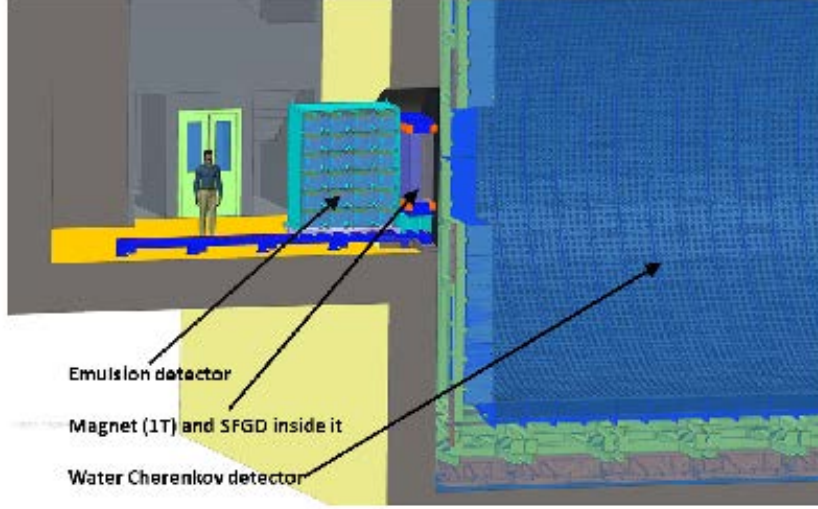


Figure 1: The three detector components of the near detector complex, the emulsion detector, the SFGD and the water Cherenkov detector. The direction of the neutrino beam is left to right, perpendicularly to the face of the WC tank. Illustration from [4].

Lund, Sweden [3] can be modified to provide such a high intensity neutrino beam, and the ESS ν SB is the experiment proposed to measure CP violation using this beam [4, 5]. ESS ν SB will measure the flavour composition of the unoscillated neutrino beam immediately after the beam target at the near detector, which consists of a water Cherenkov (WC) detector, a magnetized scintillator detector called the Super Fine-grained Detector (SFGD), and an emulsion-based detector similar to the NINJA detector employed at J-PARC [6]. The components of the near detector are illustrated in Figure 1. The flavour composition of the oscillated neutrino beam will be measured at the far detector, composed of two Cherenkov tanks located 340 km from the target such that the flux will contain both the first and second maximum of muon (anti)neutrino to electron (anti)neutrino oscillations [4]. By measuring the flavour composition of both the neutrino and antineutrino beam the oscillation probability can be calculated and δ_{CP} can be estimated.

To enable these measurements, accurate reconstruction of the neutrino events is a necessity. While accurate reconstruction methods exist and are implemented in neutrino experiments [7, 8], they are often based on likelihood methods and have limits to their performance and reconstruction speed. The latter becomes particularly important in the development stages of the ESS ν SB, as having fast event reconstruction methods enable faster exploration of different detector designs and PMT layouts. Recently, Machine Learning (ML) methods and particularly methods based on Graph Neural Networks (GNNs) have proven useful for neutrino event reconstruction [9]. In the following sections, we explore the possibilities of replacing the current likelihood-based method with one based on a GNN for classification of flavour and interaction type.

2 Methods

2.1 Simulation

The GNN models described in the following sections were trained and subsequently benchmarked against the likelihood method applied to two data sets of Monte Carlo (MC) simulated events (used in the ESS ν SB conceptual design report (CDR) [4] and [10]). The events were simulated using the neutrino interaction vertex generator GENIE [11, 12] for neutrino interactions, and the water Cherenkov simulation software WCSim [13] for particle transport, and the detector response. WCSim is based on Geant4 [14] and developed by the Hyper-Kamiokande collaboration. For these simulations the Near Detector tank is configured to have a volume of 750 m³ with a photomultiplier tube (PMT) coverage of 30%.

The first dataset contains samples of single charged leptons. These events have been generated homogeneously distributed within the detector tank with an isotropic direction distribution, and uniformly covering kinetic energies from 0 GeV to 1.5 GeV. In the absence of neutrinos and neutrino interactions for this dataset, the flavour of the charged lepton was the target variable for classification.

The second dataset contains a uniform distribution of interacting neutrino energies from 0 GeV to 1.5 GeV, with pure water as target material. These events were classified by the flavour of the interacting neutrino and later by interaction type. The simulations contain only muon neutrino and electron neutrino events (muons and electrons for the charged lepton simulations), as the tau neutrinos were omitted from this study¹. Both events with charged current (CC) and neutral current (NC) interactions were considered.

2.2 Reconstruction

2.2.1 Graph Neural Networks

The approach developed for this study uses GNNs. GNNs are an ML method based on graph theory, in which data is expressed as graphs of nodes connected by edges, and node information is propagated between the layers of the network through the exchange of information with neighbouring nodes (known as message passing) [15]. In the case of this study, a neutrino interaction event (or an event containing a single charged lepton) is described by a single graph, with each PMT signal being a node with features such as PMT position, signal arrival time and charge recorded. The graph is completed by connecting the nodes with edges according to a K-nearest neighbours scheme (in x, y, z, t).

To train and apply GNN models to the graphs, the GraphNeT [16] framework is used. Initially developed by the IceCube collaboration [17, 18], GraphNeT facilitates training of ML models (including GNNs) on data structures and tasks commonly associated with Cherenkov detectors and neutrino telescopes, using the PyTorch [19], PyTorch Geometric [20] and Pytorch Lightning [21], libraries as backend. GraphNeT takes as input tables of PMT signals and event labels, and handles parallel loading of data and building of graphs during training, outputting trained models and predictions for the desired data sets. In addition to useful loss functions and other optimization

¹With a beam energy of up to 2 GeV and the tau lepton mass $m_\tau \approx 1.777$ GeV, the tau neutrinos are not produced in significant amounts at the target, and will not partake in CC interaction in the detectors.

utilities, GraphNeT contains various model architectures like the DynEdge model. The DynEdge architecture is based on the EdgeConv [22] message passing scheme, but with the addition of edge reconfiguration between each graph convolution [9], and is employed for this study without modifications.

For each event, the GNN model will output a model score between 0 and 1, representing the probability assigned by the model that the event is of the signal flavour, where a score closer to 1 indicates higher probability of signal flavour. The selection threshold can be adjusted within the this range to obtain a satisfactory sample. For this study, the default signal flavour will be electron flavour, and deviations will be addressed when relevant.

2.2.2 Likelihood-based Reconstruction

The baseline method used for classification is based on the fitQun algorithm [8, 23]. A log-likelihood-based reconstruction method developed for the Super-Kamiokande Experiment and based on methods developed for the MiniBooNE experiment [24], it can accurately reconstruct the flavour and kinematic parameters of the interaction products, from which information about their neutrino parents can be inferred.

For a given event, the fitQun framework will provide a negative logarithmic likelihood (NLL) estimation for each relevant lepton flavour hypothesis. The model score in this work follows the approach of the CDR [4] and is taken to be the ratio between the NLL of an event being a muon (neutrino) event and the NLL of being an electron (neutrino) event.

$$\text{Model score} = \frac{NLL^{\text{fitQun}_\mu}}{NLL^{\text{fitQun}_e}} \quad (2.1)$$

2.2.3 Training

Both data sets (neutrinos and charged leptons) were classified by neutrino or lepton flavour (and later interaction type and pion production) using the same GNN models, which were trained only on the neutrino event data set. Training separate models applied to the charged lepton events was investigated, but did not yield any significant increase in performance. The models were trained on 100.000 (67% training, 33% validation) events on two NVIDIA Tesla K20 cards and evaluated on ~50.000-300.000 events (depending on the cuts applied and the resulting sample sizes). For a more detailed overview of training and test samples, see Table 1. The training was carried out using a linear 1cycle learning rate scheduler [25] with early stopping, with all models converging within 50 epochs.

2.2.4 Performance Metric

The neutrino beam produced at the ESS ν SB target will contain > 98% muon neutrinos [4] with the remainder dominated by electron-neutrino. The main purpose of the Near Detector (on which this study is focused) is to accurately determine the electron neutrino content of the initial neutrino beam (before oscillations) in order to reduce systematic uncertainties for the oscillation analysis conducted with far detector data. For these reasons, the purity of the electron neutrino sample (which is of greater importance) is more severely affected by the misidentification of muon neutrino events, making it more important to filter out muon neutrino events from the electron neutrino sample

	Simulated	After cuts	Training	Test w/ tcuts	Test w/o cuts
Charged leptons	100,000	49,325	-	49,325	-
Neutrinos (CC)	400,000	226,425	100,000	100,000	300,000

Table 1: Breakdown of the sizes of training and test samples for the charged lepton and CC neutrino interaction data sets. The charged lepton data set was not used for training, and was not tested without cuts. All samples were simulated with a 1:1 ratio of initial lepton/neutrino flavour, and a 1:1 ratio of ν and $\bar{\nu}$ for the neutrino sample. The training was divided into 67% training data and 33% validation data with the same ratios.

For the neutrino events, a total of 800,000 events of equal parts CC and NC events were simulated, but as the NC events are disregarded for the majority of this study, they are not covered here.

		Prediction				Prediction	
		ν_μ	ν_e			ν_e	ν_μ
Truth	ν_μ	True Positive (TP)	False Negative (FN)	Truth	ν_e	TP	FN
	ν_e	False Positive (FP)	True Negative (TN)		ν_μ	FP	TN
(a) ν_μ signal				(b) ν_e signal			

Table 2: Confusion matrices for flavour classification with the signal flavours ν_μ (a) and ν_e (b).

than electron neutrinos from the muon neutrino sample. Thus, this work follows the example of the CDR [4] and [10] and uses a false positive rate (FPR) of 0.1% for (muons neutrinos identified as) electron neutrinos and 1% for (electron neutrinos identified as) muon neutrinos as the target parameter when optimizing the selection criteria. The true positive rate (TPR) is the same as the efficiency, while the FPR is the probability of labelling an event of the background flavour as signal, quantified e.g. for electron neutrinos, as the number of muon neutrino events identified as electron neutrino events relative to the total number of muon neutrinos events, and is thus independent of the ratio between the number of events of each flavour. Purity could also be used for this study, as a pure electron neutrino sample is the key objective, but since we exactly want to measure the beam composition because it is not known precisely and it is not reflected in the simulations, the TPR and FPR are better suited. For illustration, assuming a 90% TPR (and 98% ν_μ , 2% ν_e), the 0.1% target TPR will yield an electron neutrino sample of 95% purity, demonstrated in Eq.2.2. Extended definitions of true and false positive and negative and their dependence on the signal flavour for flavour classification are illustrated in the confusion matrices in Table 2.

$$n_{\nu_\mu} = 98\% \cdot 0.1\% = 0.098\% \quad n_{\nu_e} = 2\% \cdot 90\% = 1.8\% \quad \text{purity}_{\nu_e} = \frac{1.8\%}{1.8\% + 0.098\%} = 95\% \quad (2.2)$$

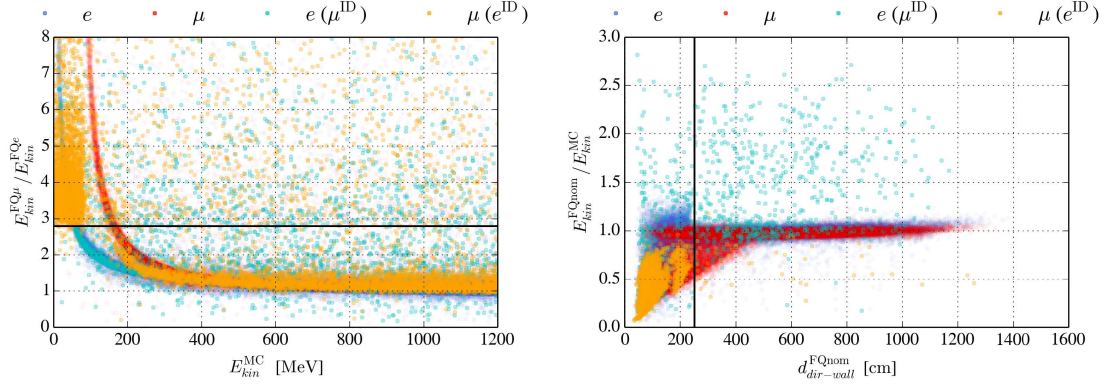


Figure 2: Illustration of two of the likelihood-based cuts used to remove incorrectly identified events from the charged lepton data, showing correctly identified electrons (dark blue) and muons (red) and incorrectly identified electrons ($e(\mu^{ID})$, cyan) and muons ($\mu(e^{ID})$, orange) distributed by reconstructed variables (and in one case, simulated energy). The classification uses the fitQun-based approach described in [4].

The sub-Cherenkov criterion (left) shows events distributed by simulated energy (E_{kin}^{MC}) and ratio between reconstructed energies using a muon ($E_{kin}^{FQ\mu}$) and electron (E_{kin}^{FQe}) hypothesis, respectively. The black line indicates the cut, (removing events with $E_{kin}^{FQ\mu} / E_{kin}^{FQe} > 2.8$), which serves to reject muon events where the muon has too little energy to produce Cherenkov radiation, and as a result is primarily detected through its electron decay product and misidentified an electron.

The Cherenkov-ring resolution criterion (right) show events distributed by distance from the interaction vertex to the detector wall along the direction of the charged lepton d_{wall}^{FQnom} , and the ration the reconstructed and simulated energies. The black line indicates the cut, removing events with $d_{wall}^{FQnom} < 250$ cm which are often misidentified.

Figures from [4].

2.3 Likelihood-based Cuts

For the reconstructions in the CDR [4], for which the likelihood-based method was used, a number of cuts² are applied to the data to filter out events that are hard to reconstruct with good accuracy. These cuts are based on reconstructed variables from the fitQun-based reconstruction and include, as example, the cuts shown in Figure 2. On the left, the ratio between reconstructed energies using a muon (neutrino) and electron (neutrino) hypothesis, respectively, is used to filter out muon (neutrino) events where the produced muons have energy below the Cherenkov threshold. These events are primarily detected via the electron from the muon decay, and thus frequently misidentified as electron (neutrino) events. On the right, the distance from the starting initial position of the charged lepton (which would be the interaction vertex) and the detector wall along the direction of travel of the lepton, is used to filter out events where the charged lepton is produced to close to the wall, which are often misidentified.

As these cuts are based on variables from the fitQun-based reconstruction, replacing the

²For a full list of cuts, see [4], Sec. 7.2.3

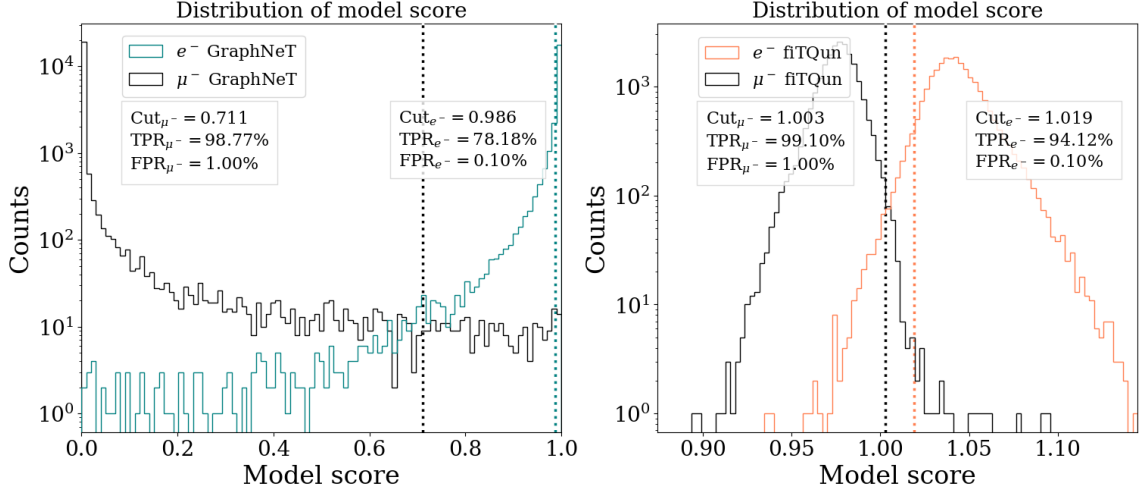


Figure 3: Model score distributions for the GNN (left) and fitQun (right) lepton flavour classification method applied to the charged lepton events (with likelihood-based cuts). Dashed lines represent thresholds corresponding to target FPRs, and the resulting TPR is shown.

likelihood-based reconstruction method with a GNN-based method also means losing the ability to perform these cuts. Additionally, about half of the events are lost when applying these cuts, so removing or loosening the cuts could provide better statistics. Thus, part of this work is also investigating if the GNN classification model can perform well without these likelihood-based cuts applied.

3 Results

3.1 Charged Lepton Events

For charged lepton event simulations, the likelihood-based cuts described in 2.3 were applied to the data set to have a baseline case to compare both with the CDR results and subsequent studies in this paper. The resulting sample should have a large fraction of unambiguous events that can be easily classified by both reconstruction methods. The model score distribution of the charged lepton event classification is shown in Figure 3 for the GNN (left) and fitQun (right). Both the fitQun and GraphNeT classification methods produce acceptable results on par with the performance described in the CDR, with the fitQun-based method having slightly better performance, yielding TPRs of 99.10% and 94.12% for the chosen FPR rates for muons and electrons, respectively, against the 98.77% and 78.18% TPRs of the GraphNeT-based method. This is to be expected, since this is the type of events the likelihood-based method is optimized for.

Figure 4 shows Receiver Operating Characteristic (ROC) curves for both reconstruction method. ROC curves are obtained by varying the model score threshold and plotting the obtained FPR (x-axis) and TPR (y-axis) and thus illustrate the trade-off between the FPR and TPR for a binary classification model. The shape of the ROC curve with the electron flavour as signal (Figure 4, left) indicates that for looser constraints on FPR ($> 5 \cdot 10^{-4}$) the fitQun-based method yields greater TPRs than the GNN, while for lower on FPRs, the GNN starts to outperform the fitQun-based

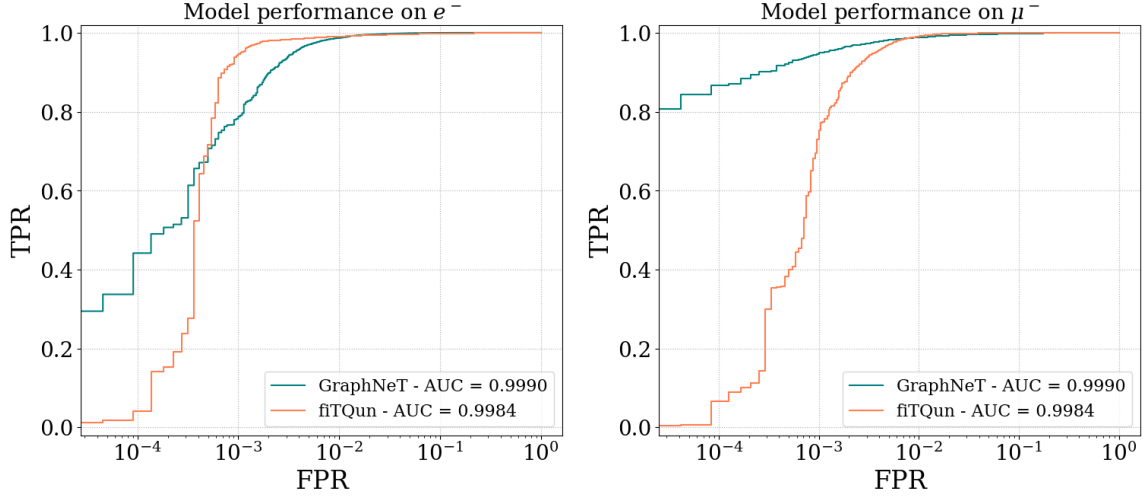


Figure 4: ROC curves for the GNN and fiTQun lepton flavour classification methods applied to the charged lepton events (with likelihood-based cuts) with electrons as signal (left) and muons as signal (right).

method. For muons (Figure 4, right) the performance of the two methods is similar above $\text{FPR} = 10^{-2}$, while for lower target rates, the GNN has superior performance. Figure 4 also shows the calculated area under the ROC curves (AUC). The AUC describes a classification method's ability to yield a high TPR at the full FPR range, with a perfect ROC curve being a 90° angle (or a horizontal line when the FPR-axis is logarithmic) with $\text{AUC} = 1$. The AUC is useful as a single statistic to compare the performance of reconstruction methods, but for more rigorous testing one should focus on the specific FPR requirements.

3.2 Charged Current Neutrino Events

3.2.1 With Likelihood-based Cuts

In this section the performance of the GNN and fiTQun-based classification methods applied to full neutrino interaction event simulations with the likelihood-based cuts described in 2.3 applied is shown, in order to illustrate how each method performs on neutrino simulations under optimal conditions.

The model score for the neutrino event classification distributed by true (simulated) neutrino energy is shown in Figure 5 for the GNN (left) and fiTQun (right). The GNN-based classification yields TPRs of 96.98% and 69.65% for muon and electron neutrinos respectively, when applying the same restrictions on FPR, which is comparable to the performance of the reconstruction described in the CDR [4]. The likelihood-based classification yields a similar TPR of 92.86% for muon neutrinos, but a significantly lower TPR of 0.53% for electron neutrinos. Although the two groups seem well separated by the likelihood-based method with no clear structure to the misidentified events, the 0.1% FPR for electron neutrinos is too tight to produce a useful TPR.

Figure 6 shows ROC curves for both reconstruction methods, confirming that the GNN consistently outperforms the fiTQun-based method for all relevant FPR rates, and that the performance difference is the greatest for lower ($10^{-2} - 10^{-4}$) FPR ranges of electron neutrino events. This

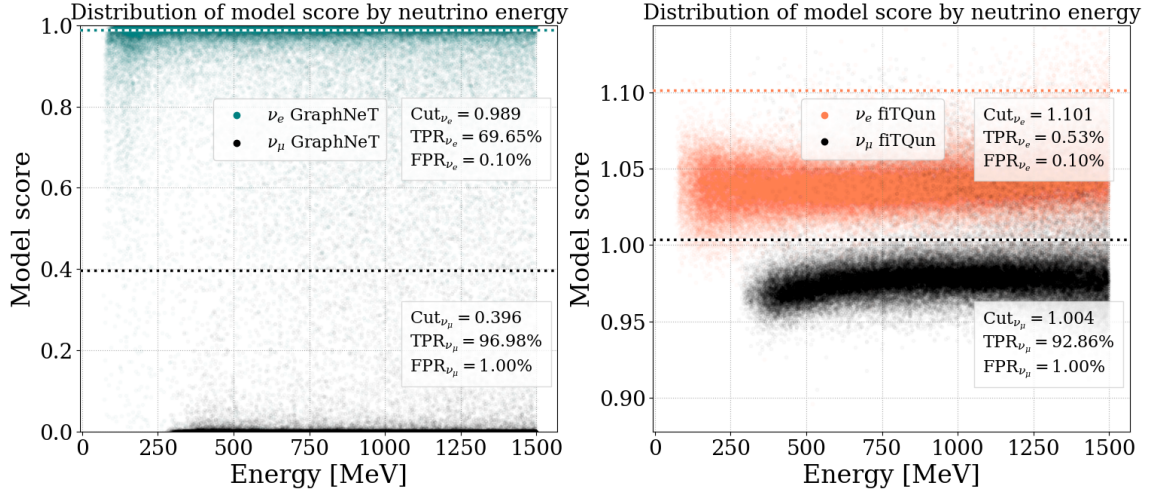


Figure 5: Model score distributed by true (simulated) neutrino energy for the GNN (left) and fiTQun (right) neutrino flavour classification methods applied to CC neutrino events with likelihood-based cuts applied. Dashed lines represent thresholds corresponding to target FPRs, and the resulting TPR is shown.

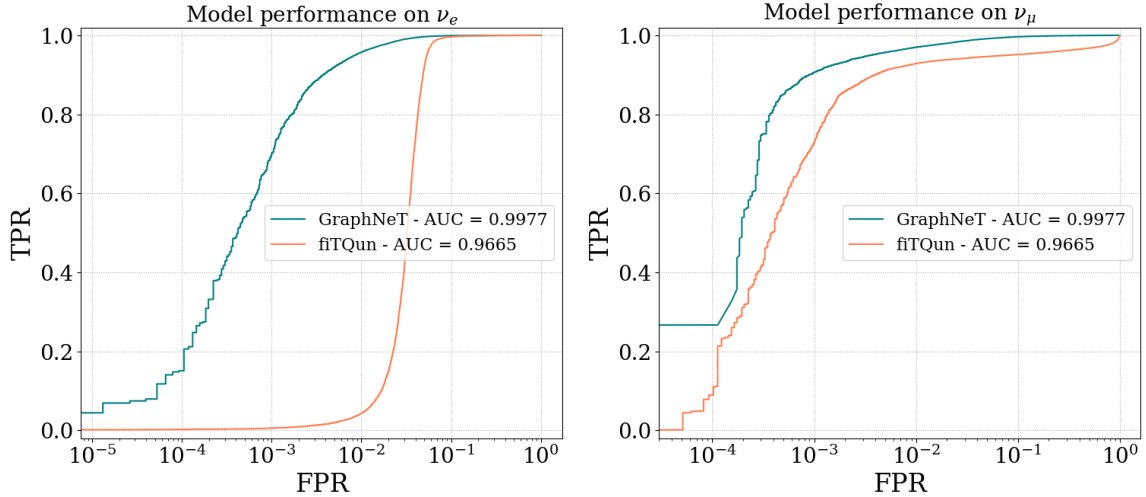
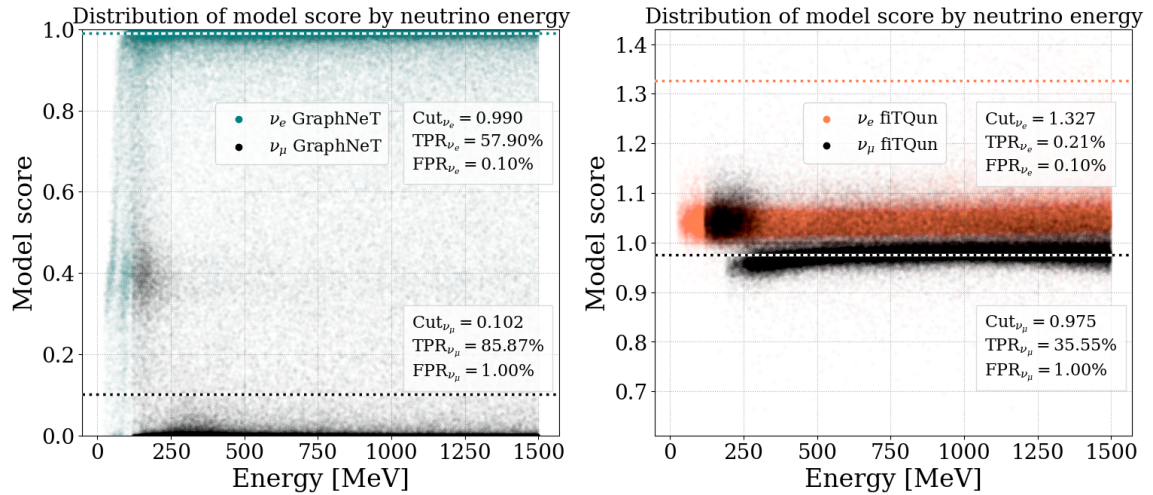
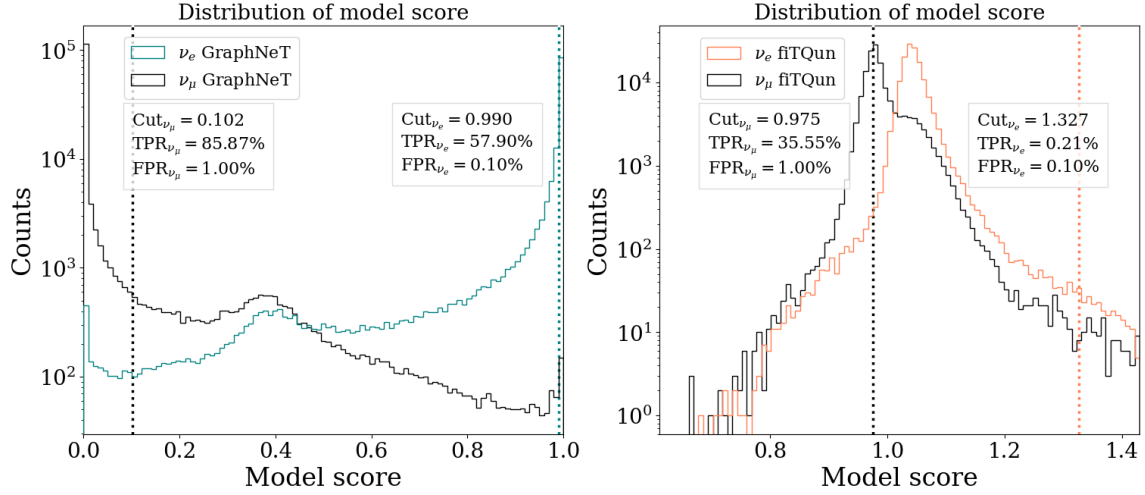


Figure 6: ROC curves for the GNN and fiTQun neutrino flavour classification methods applied to CC neutrino events with likelihood-based cuts applied with electron neutrinos as signal (left) and muon neutrinos as signal (right). Similarly to the previous data sets, the ROC curves become jagged in the low FPR region, due to low statistics, but notably the ROC curve for the GNN on ν_μ becomes flat at around 0.01% FPR, indicating the about 25% of the muon neutrinos receive such a high model score that all the electron neutrinos can be filtered out without losing this fraction of muon neutrinos.

suggests that there is a large group of muon neutrino events that are falsely identified as electron neutrinos by fiTQun, but that the GNN can correctly identify. We return to this point in Section 3.2.3.



3.2.2 Without Likelihood-based Cuts

In the following, the likelihood-based cuts described in 2.3 were disregarded to investigate if the GNNs can be applied to the data without these cuts, which as argued before is important if the GNNs are to be used to optimise the detector geometry. The model score distributions of the neutrino event classification is shown in Figure 7 for the GNN (left) and fitQun (right). For this sample, the performance of both methods is similar to the performance with the likelihood-based cuts applied, though it is worse in all cases. The GNN classification yields TPRs of 85.87% and 57.90% for

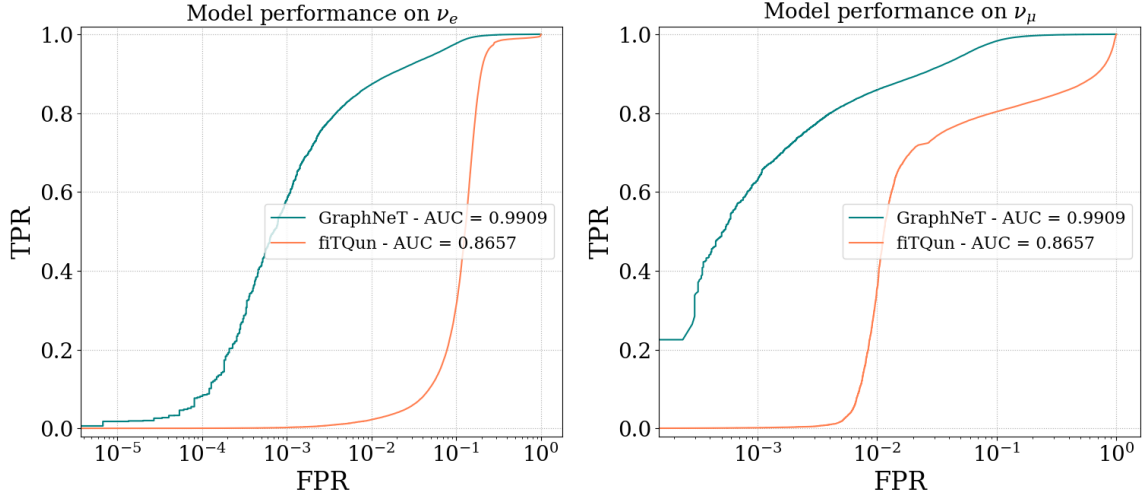


Figure 9: ROC curves for the GNN and fiTQun neutrino flavour classification methods applied to CC neutrino events without likelihood-based cuts applied with electron neutrinos as signal (left) and muon neutrinos as signal (right).

muon and electron neutrinos respectively, and the likelihood-based classification yields lower TPRs of 35.55% and 0.21% for muon and electron neutrinos respectively.

The model score distributed by true (simulated) neutrino energy shown in Figures 8 for the GNN (left) and fiTQun (right), shows a clear band of muon neutrinos in the lowest energy range, that overlaps completely with the electron neutrinos, making it hard to exclude them without losing the majority of electron neutrinos. Since muons produced in interactions in this range will have low enough energy that the dominant part of their event signature will be from the electrons produced from their decay, such a band is expected. While the GNN performance also appears to be worst in the low energy range, such a band is not present. Instead, many of the ambiguous events obtain a model score of about 0.4, leading to them being excluded when applying the restrictions on FPR. This indicates that the GNN, from being exposed to enough of these events during training, can learn to identify them as ambiguous and assign them a model score that represents this.

Figure 9 shows ROC curves for both methods, confirming again that the GNN consistently outperforms the fiTQun-based method, and that while the difference in performance is still greatest for lower FPR electron neutrino events, the performance difference for muon neutrino events is greater than with the likelihood-based cuts applied.

3.2.3 Performance Comparison

Many factors impact the performance of the the reconstruction methods applied to the complex neutrino interaction events. In the following, a few examples of events where the two methods have similar or dissimilar performance are inspected. While the examples do not exhaust all the relevant event topologies, they can provide valuable insight about features to consider when designing the ESS ν SB reconstruction pipeline. Figure 10 shows event displays in 3D (top) and 2D (bottom) of selected neutrino events, created to investigate the difference in performance on neutrino data between the two methods. The events were selected by the difference between the model score and

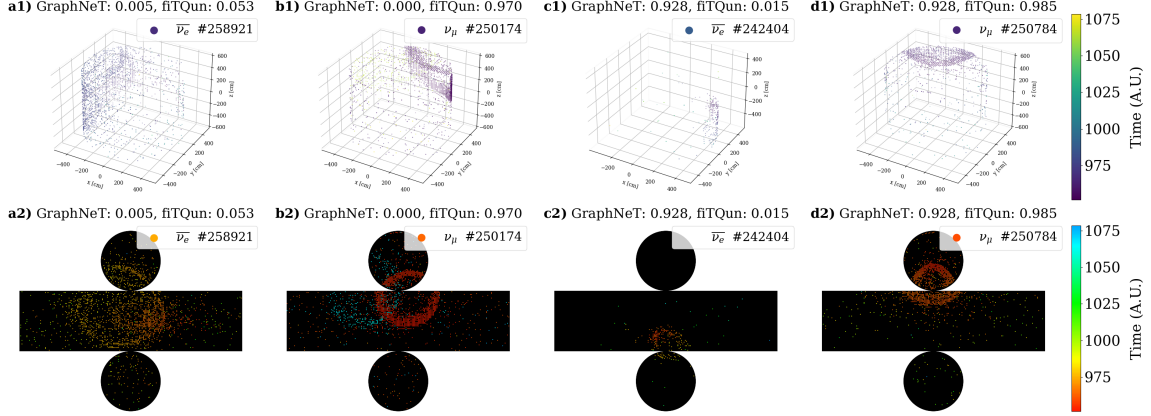


Figure 10: 3D (top) and 2D (bottom) event displays of events with different combinations of Confidence-Correctness Scores for each neutrino flavour classification method as explained in the text. The event displays show PMT signals distributed by the spatial (x, y, z) position of the PMT, with the colour indicating the arrival time of the signal.

the truth label (0 for ν_μ , 1 for ν_e), here denoted the Confidence-Correctness Score (CCS), for each method. The CCS, illustrated in Eq. 3.1, is a measure of how "mistaken" the classification method is for a given event, regardless of the flavour of the event. A CCS of ≈ 0 (a correctly confident classification) can either be a ν_μ event (0) with a model score of ≈ 0 , or a ν_e event (1) with a model score of ≈ 1 . In contrast, a CCS of ≈ 1 signifies a confidently incorrect prediction, and arises from a model score of ≈ 0 for a ν_e event (1) or vice versa. In the following, the events were selected to have different combinations of CCS's for the two reconstruction methods³.

$$CCS = | \text{truth} - \text{prediction} | \quad (3.1)$$

The first event from the left (Figure 10, a1 and a2) shows an event where both methods predict correctly with great confidence. This event has many detector hits, providing lots of information to the reconstruction algorithms, as well as a clear ring structure. The second event (Figure 10, b1 and b2), where the GNN-based method predicts correctly and the fiTQun-based method incorrectly, is of particular interest, as both the 2D and 3D event displays show two distinct Cherenkov Rings with a significant time delay, most likely from an event where a muon neutrino produces a muon that decays and produces an electron. This indicates that the GNN has learned to identify these more complex events out of the box, even for cases where it is difficult for fiTQun, a finding that is consistent when looking at Figure 19 of Appendix A, which shows several event displays of similar CCS combination, many of which exhibit double rings. The third event (Figure 10, c1 and c2) where the fiTQun-based method predicts correctly with great confidence, and the GNN predicts incorrectly, has few detector hits, but a clear ring structure. This can be interpreted as the event not having enough data points for the GNN - which as an ML model needs more information to function

³Combinations of: low-low, low-high, high-low, high-high, where a low score ($CCS < 0.1$) indicates a correct prediction with high confidence, and a high score ($CCS > 0.9$) indicates an incorrect prediction with high confidence. The value 0.1 was chosen to accept most of the unambiguous events and the value $0.9 = 1 - 0.1$ is derived from that.

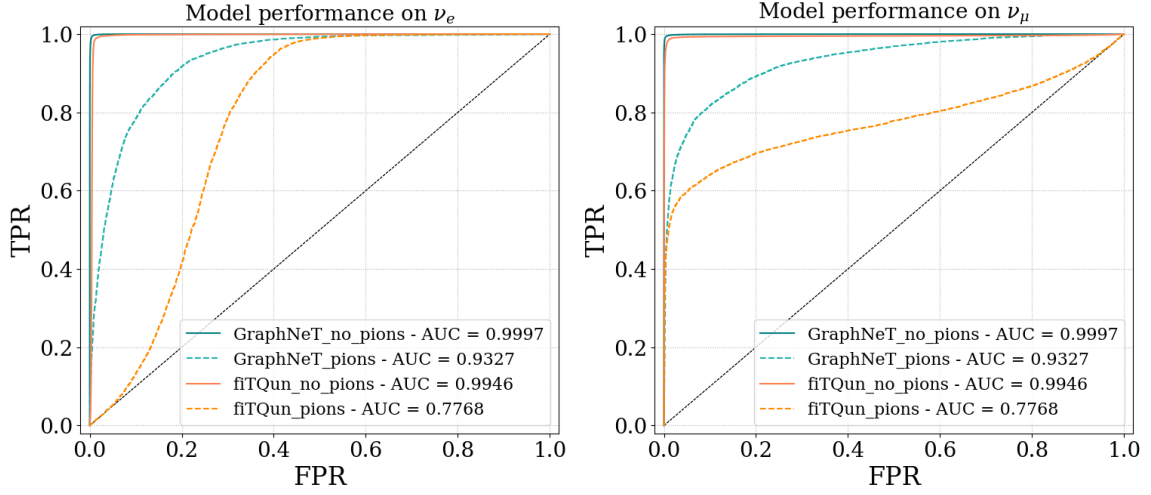


Figure 11: ROC curves for the GNN and fitQun neutrino flavour classification methods applied to CC neutrino events (with likelihood-based cuts) with electron neutrinos as signal (left) and muon neutrinos as signal (right) in the cases where pions and no pions are produced in the final state.

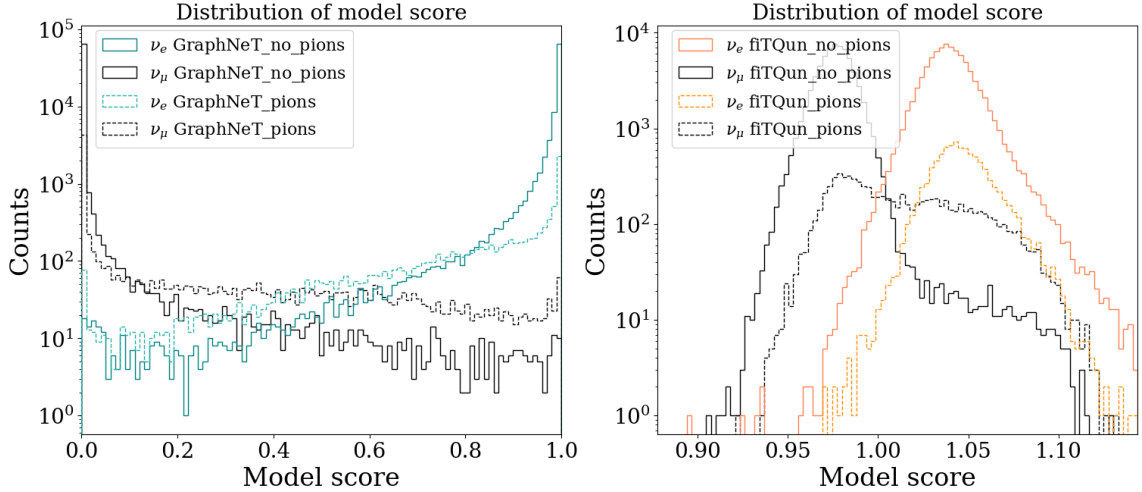


Figure 12: Model score distributions for the GNN (left) and fitQun (right) neutrino flavour classification methods applied to CC neutrino events (with likelihood-based cuts) in the cases where pions and no pions are produced in the final state.

optimally - to identify the pattern, but ring-like enough for the fitQun-based method - which as an LLH-based method does not necessarily have the same limitations - to converge. The final event (Figure 10, d1 and d2) has an average amount of detector hits, and a clear ring shape, and should be easily interpretable by both methods. It is not clear why both methods have poor performance on such an event, and future studies should look into this behaviour. Since the main focus of this section is on the events where the methods differ in performance, it is not investigated further here.

3.2.4 Pion Production

Another factor that blurs the picture of neutrino interactions in the ESS ν SB detector is the production of pions. Pions are a common product of weak neutrino-nucleus interactions, created both in resonance production processes and coherent scattering. Both CC and NC coherent scattering can produce pions, with CC processing being the dominant contributor. The pion production can interfere with measurements, both because they can create a signature that looks like Cherenkov radiation from a charged lepton, and because their energy will be missing from the charged lepton energy reconstruction and consequently from the neutrino energy estimation [26, 27].

Figure 11 shows the ROC curves for the GNN and fiTQun-based methods applied to the neutrino events (in this case with the likelihood-based cuts, to better visualise the differences in performance between the two methods), but separated into events where pions are produced in the final state (which make up $\sim 10\%$ of the events) and events where no pions are produced. It is clear, that both methods are able to classify events with no pion production, whereas for the event with pion production, the fiTQun-based method performs significantly worse than the GNN for both electron and muon neutrinos. This is also demonstrated in Figure 12 which show histograms of the model scores for the GNN (left) and fiTQun (right) classification methods. For the GNN, there is still a good separation of the events with pions, but the number of events receiving a score near 0 or near 1 is lower, while the number of events receiving scores in the rest of the spectrum is higher, signifying that the events are harder to classify. For the fiTQun-based method, the distribution of ν_e events is unchanged, but the distribution of ν_μ events are shifted to the right (higher model scores), resulting in a large overlap of ν_e and ν_μ events. This indicates, that one might want to start by using classification model (like another GNN) to separate events into classes with or without pions before running the reconstruction, in order to treat the two cases separately. We will test this idea in the next subsection.

3.2.5 Pion Classifier

Following the results in Section 3.2.4, a GNN classification model was trained on the CC (CC) neutrino interaction data set to distinguish between events with and without pion production. The model target variable was chosen to be a boolean that had the value true (1) for events where any number of pions were created in the interaction, and false (0) for events where no pions were produced. Neutrino flavour was not taken into account for this model, and the likelihood-based cuts were not applied, as the goal was to obtain the most general model possible to separate all types of events by pion production. The model score distribution and the ROC curve for the resulting model, are shown in Figure 13 on the left and right side, respectively. Together, the figures show that while the model can classify most events correctly, it is harder for the model to confidently identify events without pion production than it is for events with pion production. This can be seen from the model score distribution on the left, as almost all the events with pion production are assigned a model score close to 1, whereas the events without pion production receive a wide range of model scores. Consequently, for a FPR requirement of 1.00%, the model can retain 47.4% of the sample of events with pions. For events without pion production, the majority of the sample is lost when enforcing an FPR of 1.00%. Instead, Figure 13 shows the result of raising the requirement to 5.00% FPR,

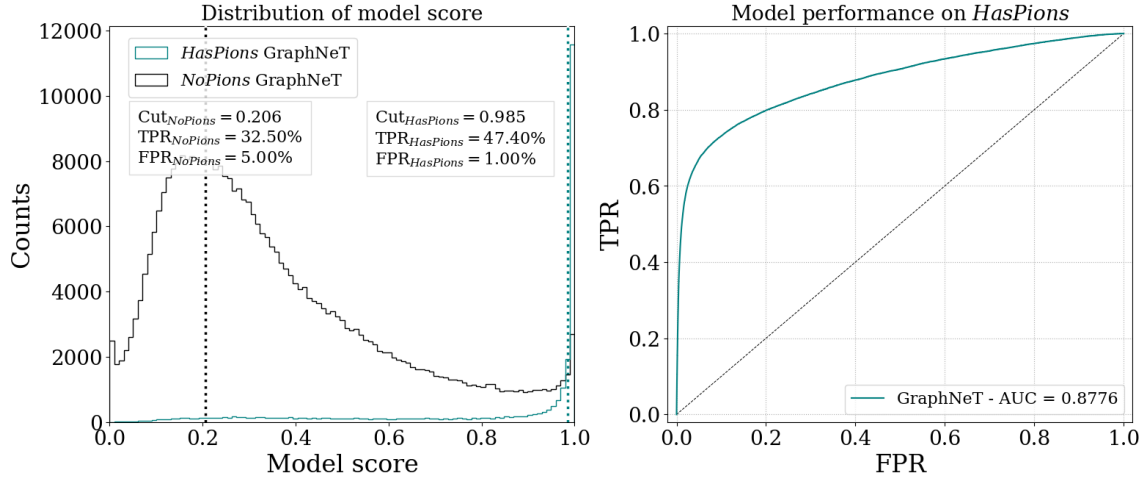


Figure 13: Model score distribution (left) and ROC curve (right) for the GNN pion production classification methods applied to CC neutrino events. Dashed lines on the left represent cuts corresponding to chosen FPRs, and the resulting TPR is shown.

which produces allow a TPR of 32.5%. The impact of using this classifier will be investigated in the following section.

3.2.6 Flavour Classification Performance Split by the Pion Model

With the pion classifier described in the previous section completed, the next objective was to study how the reconstruction for both the GNN and the fitQun-based method was impacted by filtering the data with this classifier. For this, both the GNN and fitQun-based flavour classification methods were applied to the neutrino events (in this case with the likelihood-based cuts, to better visualise the differences in performance between the two methods), but separated by the pion production classification model described in Section 3.2.5, using the thresholds shown on the left of Figure 13, to obtain reasonably pure samples of events with and without pion production. Figure 14 shows the ROC curves for two models and the two data samples, which are almost identical to the ones shown in Figure 11, showing that both methods perform very well on the events without pion production, and that the GNN has significantly better performance on the events with pion production than the fitQun-based method. This is confirmed when looking at Figures 15 which shows the model score distributions for the GNN (left) and the fitQun-based method (right). For the GNN, while the overlap is greater for the events with pion production than for the events without, there is still a significant separation for both samples. For the fitQun-based method, the separation of events without pions is very clear, whereas there is a large overlap between the two flavours for the events with pion production. This indicates that the pion production classifier is working as intended, separating events into categories with different reconstruction difficulty.

3.3 Neutral Current Neutrino Events

For the next section, the neutrino interaction dataset with both CC and NC events is used without any likelihood-based cuts. This includes 400.000 simulated NC events in addition to the CC events

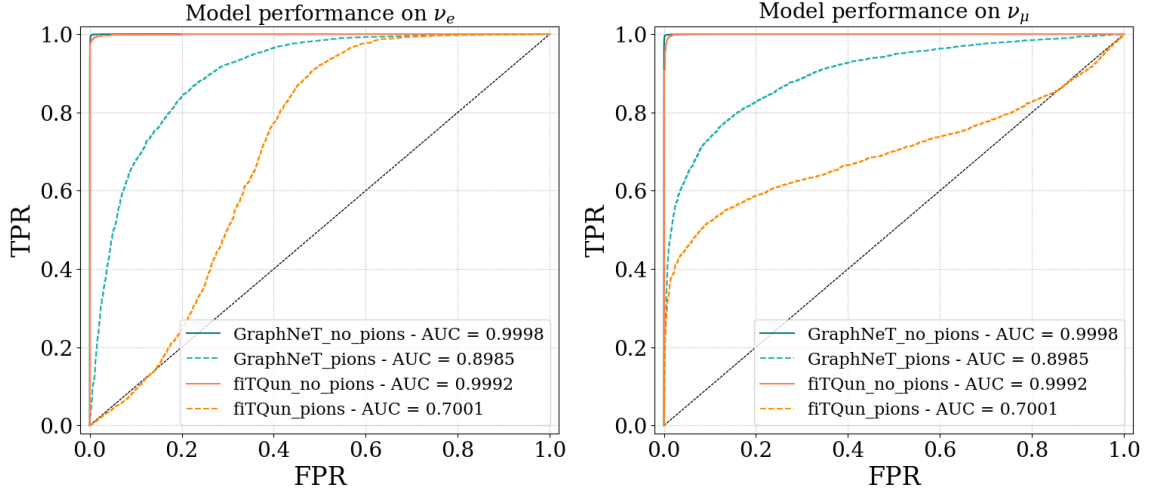


Figure 14: ROC curves for the GNN and fitQun neutrino flavour classification methods applied to CC neutrino events (with likelihood-based cuts) with electron neutrinos as signal (left) and muon neutrinos as signal (right) separated by events with and without pion production using the pion production classification model.

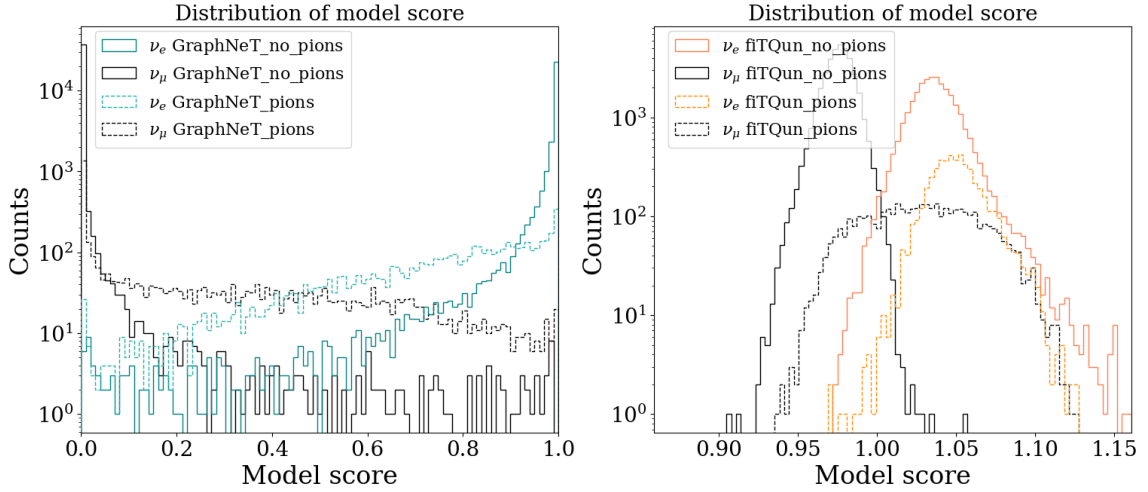


Figure 15: Model score distributions for the for the GNN (left) and fitQun (right) neutrino flavour classification method applied to CC neutrino events (with likelihood-based cuts) separated by events with and without pion production using the pion production classification model.

introduced in Table 1 and used in the previous sections. A GNN neutrino flavour classification model was trained on 100,000 events with equal representation of interaction types, flavours and neutrinos/antineutrinos, and tested on the remaining events. Figure 16 shows the ROC curves for this GNN and the fitQun-based method. The data for the ROC curves is split by CC and NC interactions, and Figure 16 shows clearly, that for the NC dataset both methods have poor performance comparable to a random selection. This is also demonstrated in Figures 17 which shows histograms of the model scores for the GNN (left) and fitQun (right) classification methods,

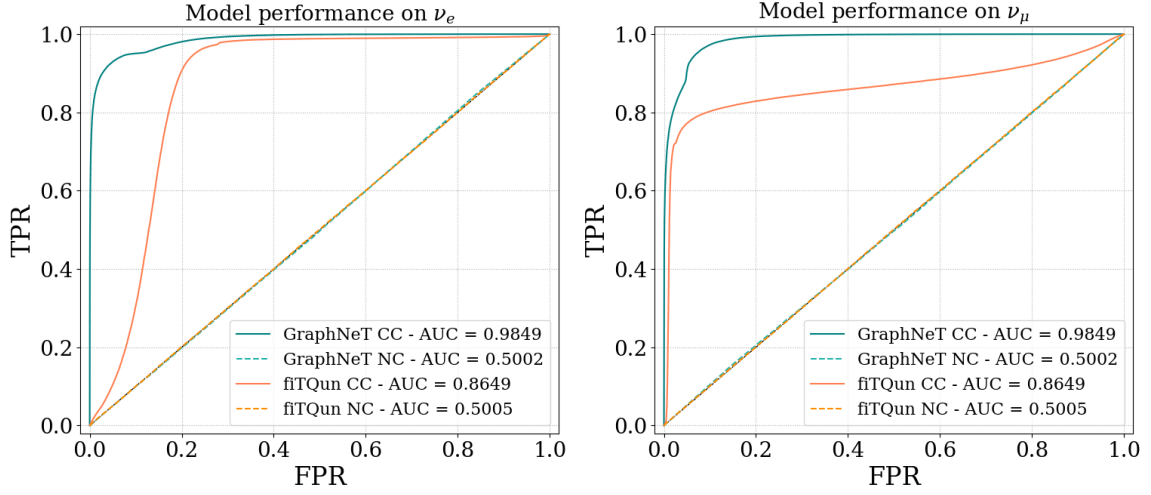


Figure 16: ROC curves for the GNN and fitQun neutrino flavour classification methods applied to CC and NC neutrino events with electron neutrinos as signal (left) and muon neutrinos as signal (right) separated by interaction type.

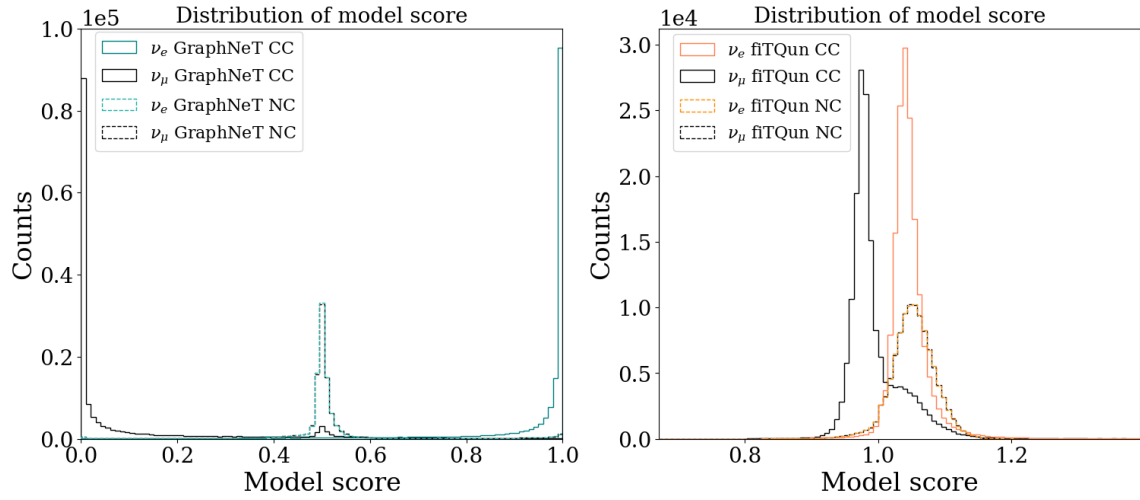


Figure 17: Model score distribution for the GNN (left) and fitQun (right) neutrino flavour classification method applied to CC and NC neutrino events separated by interaction type.

with the same data separation. For both methods there is a complete overlap in model scores for the NC events. An interesting observation is that the GNN scores almost all NC events close to 0.5, indicating that the model is unable to learn patterns that separate the events. This signifies that the development of a method for filtering out the NC events and only analyse the CC events could be beneficial.

3.3.1 Separation of CC and NC events

In an effort to meet the needs described earlier in this section, a GNN model for classifying events according to the interaction type was trained on the neutrino interaction data set with both interaction

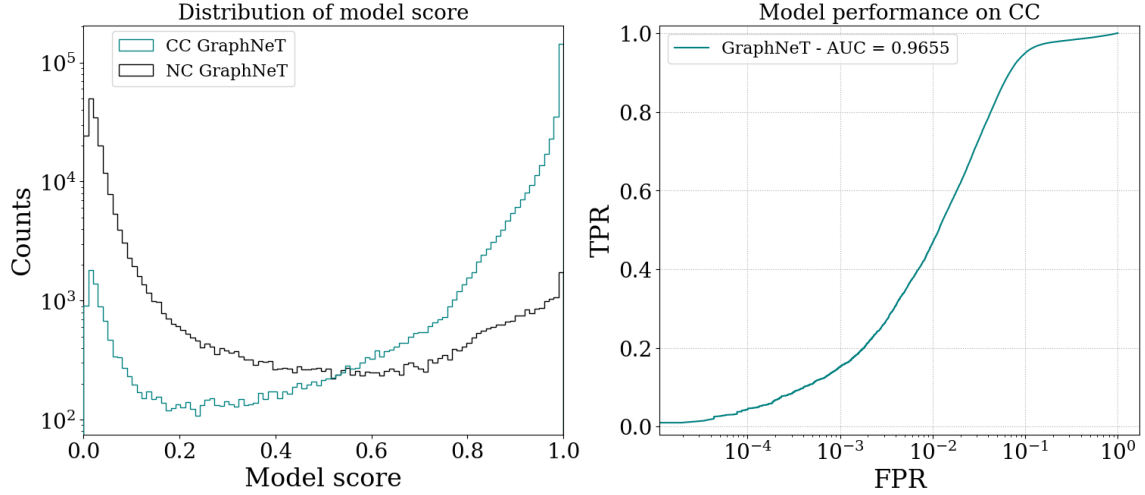


Figure 18: Model score distribution (left) and ROC curve (right) for the GNN interaction type classification method applied to CC and NC neutrino events.

types represented, and the results are shown in Figure 18. As this is not a feature of the fitQun-based method, no comparison is shown. The ROC curve in the right plot of Figure 18 shows promising performance, as 99% of NC events can be discarded while keeping a little less than half of the CC events. The histogram on the left of Figure 18 shows a greater spread of the model score for the NC events, as well as a portion of CC events being wrongly identified. But if the only goal is a pure sample of CC events, this is irrelevant.

3.4 Reconstruction Speed

Recorded training and reconstruction times for the GNN are shown in Table 3 and compared to the average event reconstruction time of the fitQun framework. The benchmarking was done using the full neutrino data set of 800,000 events.

Using the GraphNeT framework involves some overhead when converting data to one of the database formats, for which the framework is optimised. But the majority of the time consumption lies in the training step. Including model training, the total reconstruction is faster than the fitQun reconstruction by a factor of 10^3 . This trend would only improve with more events, since one only needs to train on a subset of the events, the size of which can be kept constant.

In a scenario where one could use an already trained model to reconstruct events (for instance if one makes new simulations with only minor changes to the detector layout), the reconstruction speed improvement can reach a factor of 10^4 .

4 Conclusion

In this work we propose a new classification method for the ESS ν SB based on Graph Neural Networks (GNNs) and the GraphNeT framework [16]. We demonstrate that the GNN-based method has comparable performance to the fitQun method on simulated charged lepton events with

likelihood-based cuts applied to reject difficult events, and that the GNN performance surpassed that of fiTQun for simulated CC neutrino events both with and without the aforementioned cuts.

We also present results that indicate that the GNN-based method is able to accurately reconstruct more events with two Cherenkov rings (from muons decaying to electrons), and demonstrate how GNNs can be trained to separate CC and NC events as well as events with and without pion production, and how this is beneficial, as the event types differ in reconstruction performance.

The reconstruction carried out using the GNN method is demonstrated to have a factor of 10^4 increase in performance speed from the fiTQun method. This will be very beneficial to the ESS ν SB experiment, providing great flexibility in the detector design phase.

A Events Displays of Events with Superior GNN Performance

Data extraction	10^{-4} mins/event
Training	10^{-3} mins/event
Reconstruction	10^{-4} mins/event
fiTQun Reconstruction	1 min/event
Improvement	10^3 (w/ training) / 10^4 (w/o training)

Table 3: Comparison of reconstruction speeds of the fiTQun and GNN classification methods, with and without training of the GNN.

Acknowledgments

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

The authors acknowledge support provided by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska -Curie grant agreement No 860881-HIDDeN.

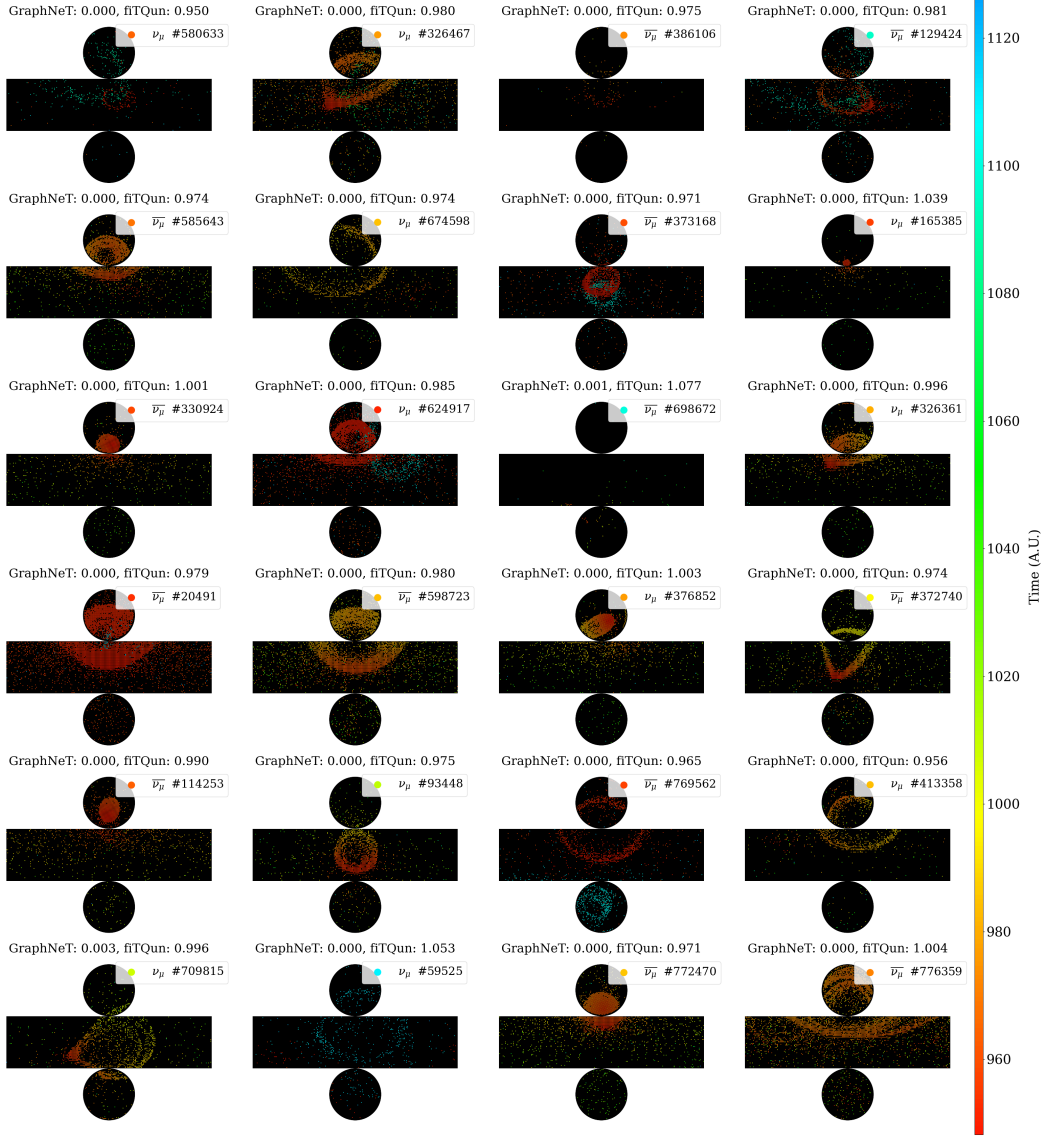


Figure 19: Events displays of randomly selected CC interaction neutrino events for which the GNN flavour classification method predicts correctly with great confidence, while the fiTQun-based method predicts incorrectly with great confidence. Several of the event displays show multiple rings of Cherenkov light, indicating that the GNN has the ability to accurately classify low energy muon neutrino events.

The training of models and reconstruction of events were performed using the GraphNeT framework (Apache 2.0), developed and maintained by GraphNeT group. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing (SNIC) at LUNARC partially funded by the Swedish Research Council through grant agreements no. 2022-06725 and no. 2018-05973.

The authors acknowledge the assistance from C. Vilela, E. O’Sullivan, H. Tanaka, B. Quilain and M. Wilking for the use of the WCSim and fitQun software packages.

References

- [1] A. Cervera, A. Donini, M. Gavela, J. Gomez Cadenas, P. Hernández, O. Mena et al., *Golden measurements at a neutrino factory*, *Nuclear Physics B* **579** (2000) 17–55.
- [2] P. Coloma and E. Fernandez-Martinez, *Optimization of neutrino oscillation facilities for large θ_{13}* , *Journal of High Energy Physics* **2012** (2012) .
- [3] R. Garoby et al., *The european spallation source design*, *Physica Scripta* **93** (2017) 014001.
- [4] ESSnUSB collaboration, *The european spallation source neutrino super-beam conceptual design report*, *The European Physical Journal Special Topics* **231** (2022) 3779–3955.
- [5] ESSnUSB collaboration, *The ESSnuSB design study: Overview and future prospects*, *Universe* **9** (2023) 347.
- [6] T. Fukuda, S. Aoki, S. Cao, N. Chikuma, Y. Fukuzawa, M. Gonin et al., *First neutrino event detection with nuclear emulsion at j-parc neutrino beamline*, *Progress of Theoretical and Experimental Physics* **2017** (2017) .
- [7] R. Abbasi et al., *Low energy event reconstruction in icecube deepcore*, *The European Physical Journal C* **82** (2022) .
- [8] T2K collaboration, *Improving the t2k oscillation analysis with fitqun: A new maximum-likelihood event reconstruction for super-kamiokande*, *Journal of Physics: Conference Series* **888** (2017) 012066.
- [9] R. Abbasi et al., *Graph neural networks for low-energy event classification & reconstruction in icecube*, *Journal of Instrumentation* **17** (2022) P11003.
- [10] A. Burgman, J. Park, J. Cederkall and P. Christiansen, *The ess neutrino super-beam near detector*, in *Proceedings of The European Physical Society Conference on High Energy Physics — PoS(EPS-HEP2021)*, EPS-HEP2021, p. 797, Sissa Medialab, Jan., 2022, DOI.
- [11] C. Andreopoulos, A. Bell, D. Bhattacharya, F. Cavanna, J. Dobson, S. Dytman et al., *The genie neutrino monte carlo generator*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **614** (2010) 87–104.
- [12] C. Andreopoulos, C. Barry, S. Dytman, H. Gallagher, T. Golan, R. Hatcher et al., *The genie neutrino monte carlo generator: Physics and user manual*, 2015.
- [13] E. O’Sullivan, “The wcsim geant4 application - github.” <https://github.com/WCSim/WCSim>, 2021.
- [14] S. Agostinelli et al., *Geant4—a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003) 250.

- [15] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner and G. Monfardini, *The graph neural network model*, [*IEEE Transactions on Neural Networks* **20** \(2009\) 61](#).
- [16] A. Sogaard et al., *Graphnet: Graph neural networks for neutrino telescope event reconstruction*, [*Journal of Open Source Software* **8** \(2023\) 4971](#).
- [17] R. Abbasi et al., *The design and performance of icecube deepcore*, [*Astroparticle Physics* **35** \(2012\) 615–624](#).
- [18] M. Aartsen et al., *The icecube neutrino observatory: instrumentation and online systems*, [*Journal of Instrumentation* **12** \(2017\) P03012–P03012](#).
- [19] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan et al., *Pytorch: An imperative style, high-performance deep learning library*, 2019.
- [20] M. Fey and J.E. Lenssen, *Fast graph representation learning with PyTorch Geometric*, in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.
- [21] W. Falcon et al., *Pytorchlightning/pytorch-lightning: 0.7.6 release*, May, 2020. 10.5281/zenodo.3828935.
- [22] Y. Wang, Y. Sun, Z. Liu, S.E. Sarma, M.M. Bronstein and J.M. Solomon, *Dynamic graph cnn for learning on point clouds*, 2019.
- [23] SUPER-KAMIOKANDE collaboration, *Atmospheric neutrino oscillation analysis with improved event reconstruction in Super-Kamiokande IV*, [*Progress of Theoretical and Experimental Physics* **2019** \(2019\) 053F01](#).
- [24] R. Patterson, E. Laird, Y. Liu, P. Meyers, I. Stancu and H. Tanaka, *The extended-track event reconstruction for minibooone*, [*Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **608** \(2009\) 206](#).
- [25] L.N. Smith and N. Topin, *Super-convergence: Very fast training of neural networks using large learning rates*, 2018.
- [26] PARTICLE DATA GROUP collaboration, *Review of particle physics*, [*Phys. Rev. D* **110** \(2024\) 030001](#).
- [27] D. Rein and L.M. Sehgal, *Neutrino-excitation of baryon resonances and single pion production*, [*Annals of Physics* **133** \(1981\) 79](#).