

NOvA Neural Networks Analysis

Twymun Safford¹

¹*Michigan State University - Department of Physics and Astronomy, GEM Fellow*

NOvA is collaboration of 180 scientists and engineers from 28 institutions which studies neutrino oscillations using the existing NuMI neutrino beam at Fermilab. The NOvA experiment is designed to search for oscillations of muon neutrinos to electron neutrinos by comparing the electron neutrino event rate measured at the Fermilab site with the electron neutrino event rate measured at a location just south of International Falls, Minnesota, 810 kilometers distant from Fermilab. If oscillations occur, the far site will see the appearance of electrons in the muon neutrino beam produced at Fermilab. In this project, I describe an application of convolutional neural network (CNN) technology to the problem of identifying neutrino particle interactions in sampling calorimeters. The goal of my project was to create, tune, and implement a cosmic rejection network.

I. INTRODUCTION AND THEORY

Neutrinos are subatomic particles produced by the decay of radioactive elements and are elementary particles that lack an electric charge. The neutrino was first postulated in December, 1930 by Wolfgang Pauli to explain the energy spectrum of beta decays, the decay of a neutron into a proton and an electron. Pauli theorized that an undetected particle was carrying away the observed difference between the energy and angular momentum of the initial and final particles. Neutrinos come in three varieties, or what is known as flavors: muon neutrinos, electron neutrinos and tau neutrinos. Scientists know that neutrinos oscillate, or change from one type to another, and have seen, for example, oscillations of muon neutrinos to tau neutrinos. But scientists have not seen muon neutrinos oscillating into electron neutrinos. Unknown factors that govern neutrino oscillations have significant implications for our understanding of the makeup of the universe via this quantum effect. This quantum effect is illustrated in Figure 1.

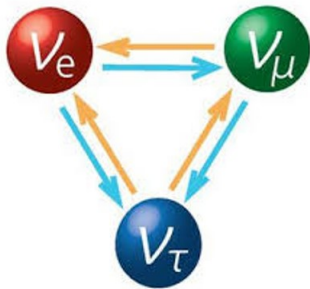


FIG. 1. This figure presents an illustration of the quantum mechanical mechanism of neutrino flavor oscillations. Presented are the three flavors of electron neutrino, muon neutrino, and tau neutrino respectively. Image courtesy of Kamioka Observatory/ICRR/University of Tokyo.

Neutrinos, once thought to be massless, are now known

to have masses that are on a scale order of million times lighter than the masses of other particles in the Standard Model of physics as depicted in Figure 2.

Drei Generationen der Materie (Fermionen)

	I	II	III	
Masse	2,3 MeV	1,275 GeV	173,07 GeV	125,9 GeV
Ladung	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{e}{p}$
Spin	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
Name	u up	c charm	t top	q e/p-Quant
				H Higgs Boson
	4,8 MeV $-\frac{1}{3}$ $\frac{1}{2}$ d down	95 MeV $-\frac{1}{3}$ $\frac{1}{2}$ s strange	4,18 GeV $-\frac{1}{3}$ $\frac{1}{2}$ b bottom	0 0 1 g Gluon
Quarks				
	<2 eV 0 $\frac{1}{2}$ ν_e Elektron-Neutrino	<0,19 MeV $\frac{1}{2}$ ν_μ Myon-Neutrino	<18,2 MeV $\frac{1}{2}$ ν_τ Tau-Neutrino	91,2 GeV 0 1 Z^0 Z Boson
	0,511 MeV -1 $\frac{1}{2}$ e Elektron	105,7 MeV -1 $\frac{1}{2}$ μ Myon	1,777 GeV -1 $\frac{1}{2}$ τ Tau	80,4 GeV ± 1 1 W^\pm W Boson
Leptonen				Eichbosonen

FIG. 2. The Standard Model of Physics which presents the mass, charge, and spin of the various subatomic particles including neutrinos. Image courtesy of HolgerFiedler nach Benutzer:Murphee via Wikimedia Commons, CC BY-SA.

The masses of the different neutrinos are unknown as is the neutrino mass hierarchy at this point in time (the mass hierarchy corresponds to which neutrino is the lightest and which neutrino is the heaviest). If it is assumed that the neutrinos ν_e , ν_μ and ν_τ , the flavor eigenstates, couple with the gauge bosons W_\pm through weak interactions are coherent superpositions of three mass eigenstates v_i , $i = 1, 2, 3$, i.e

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} * \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \quad (1)$$

where U is a unitary 3×3 matrix. In other words, it

is assumed that there are three flavor and mass neutrino states.¹ The lepton mixing matrix, U , (usually denoted as the PMNS matrix for Pontecorvo, Maki, Nakagawa and Sakata) can be parametrized in terms of the three angles Θ_{13} , Θ_{23} , and Θ_{12} and the CP phase δ_{CP} .

$$U = \begin{pmatrix} c_{13}c_{12} & c_{13}s_{12} & s_{13}e^{i\delta_{CP}} \\ c_{23}s_{12}s_{13}s_{23}c_{12}e^{i\delta_{CP}} & c_{23}c_{12}s_{13}s_{23}s_{12}e^{i\delta_{CP}} & c_{13}s_{23} \\ s_{23}s_{12}s_{13}c_{23}c_{12}e^{i\delta_{CP}} & s_{23}c_{12}s_{13}c_{23}s_{12}e^{i\delta_{CP}} & c_{13}c_{23} \end{pmatrix} \begin{pmatrix} e^{i\alpha} & 0 & 0 \\ 0 & e^{i\beta} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

where $c_{ij} = \cos \Theta_{ij}$ and $s_{ij} = \sin \Theta_{ij}$. Here, both α and β (both of which are unknown at the present) are the so-called Majorana phases that are decoupled from the phenomenon of neutrino oscillation. Figure 3 summarizes the present knowledge of neutrino masses and mixings including neutrino mixing angles, the CP phase δ_{CP} , and neutrino mass-squared differences based on the recent fit after the Neutrino 2014 conference.²³

parameter	best fit value $\pm 1\sigma$	3σ range
$\sin^2 \theta_{12}$	$0.304^{+0.012}_{-0.012}$	(0.270, 0.344)
θ_{12} (degrees)	$33.48^{+0.77}_{-0.74}$	(31.30, 35.90)
$\sin^2 \theta_{23}$	$[0.451^{+0.001}_{-0.001}]$ or $0.577^{+0.027}_{-0.035}$	(0.385, 0.644)
θ_{23} (degrees)	$[42.2^{+0.1}_{-0.1}]$ or $49.4^{+1.6}_{-2.0}$	(38.4, 53.3)
$\sin^2 \theta_{13}$	$0.0219^{+0.0010}_{-0.0011}$	(0.0188, 0.0251)
θ_{13} (degrees)	$8.52^{+0.20}_{-0.21}$	(7.87, 9.11)
δ_{CP} (degrees)	251^{+67}_{-59}	(0, 360)
$\Delta m_{21}^2 \times 10^{-5} \text{ eV}^2$	$7.50^{+0.19}_{-0.17}$	(7.03, 8.09)
(normal) $\Delta m_{31}^2 \times 10^{-3} \text{ eV}^2$	$+2.458^{+0.046}_{-0.047}$	(+2.325, +2.599)
(inverted) $\Delta m_{32}^2 \times 10^{-3} \text{ eV}^2$	$-2.448^{+0.047}_{-0.047}$	(-2.590, -2.307)

FIG. 3. Known neutrino masses, neutrino mass-squared differences, and mixing angles. Image courtesy of the 24th Interactional Conference on Neutrino Physics and Astrophysics, June 2-7th, 2014, Boston, USA. <http://neutrino2014.bu.edu/program/>

Knowledge of the mass hierarchy also will help answer the question of whether neutrinos are their own antiparticles. Particles and antiparticles have opposite charges. Because neutrinos have no electric charge, it is possible that neutrinos and antineutrinos are fundamentally the same.⁴

It is theorized that the big bang created equal amounts of matter and antimatter. When corresponding particles of matter and antimatter meet, they will annihilate one another. If the NOvA collaboration discovers that muon antineutrinos oscillate at a different rate than muon neutrinos, then it can be postulated that the symmetry between the neutrinos and antineutrinos is broken. This could point towards an origin for the asymmetry of matter and anti-matter within our universe, and thus the present state of our universe.

II. THE NOVA EXPERIMENT

The NOvA experiment consists of two functionally identical detectors in the NuMI (Neutrinos at the Main Injector) beam at Fermilab which produces a focused beam with an initial flavor composition largely dominated by ν_μ and a small intrinsic $\bar{\nu}_\mu$, ν_e , and $\bar{\nu}_e$ components. Placing the detectors off-axis at 14.6 millirads provides a narrow-band neutrino energy spectrum near 2 GeV. The Near Detector, located at Fermilab, is placed 1 km from the neutrino source; the Far Detector is located 810 km away near Ash River, Minnesota on the surface.⁵ Neutrinos are created at the source by firing protons from Fermilab's Main Injector into a graphite target. Fundamental particles come out of the collision between the protons and the target, including pions, which are charged particles. Magnetic horns are used to steer the pions in the direction they want the neutrinos to travel. The pions eventually decay into muons and muon neutrinos, which continue on the same path the pions were traveling. This neutrino beam is aimed downwards at 3.3 degrees. The beam starts out 150 feet below ground at Fermilab, and will pass six miles below the surface as it travels toward Ash River. The neutrino beam creation process is depicted in Figure 4.

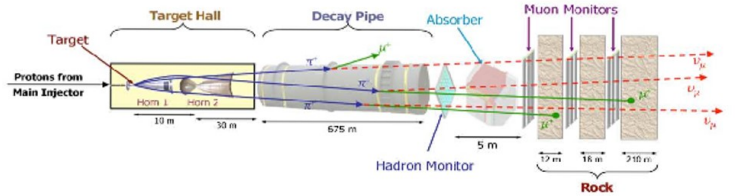


FIG. 4. A schematic of the neutrino beam configuration. Neutrinos are created by smashing protons into a graphite target which creates pions. These pions eventually decay into muons and muon neutrinos.

The NOvA detectors are composed of extruded PVC

¹ This excludes the possibility of a fourth neutrino flavor or possibly other flavors known as sterile neutrinos

² M.C. Gonzalez-Garcia, Michele Maltoni, Jordi Salvado, and Thomas Schwetz. Global fit to three neutrino mixing: critical look at present precision. JHEP, 1212:123, 2012

³ Neutrino 2014: the 24th Interactional Conference on Neutrino Physics and Astrophysics, June 2-7th, 2014, Boston, USA. <http://neutrino2014.bu.edu/program/>.

⁴ Research goals — NOvA. (n.d.). Retrieved August 5, 2018, from <http://novaexperiment.fnal.gov/research-goals/>

⁵ K. Anderson et al., The NuMI Facility Technical Design Report, FERMILAB-DESIGN-1998-01

cells filled with liquid scintillator which segment the detector into cells with a cross section 3.9 cm wide and 6.6 cm deep. These cells are 15.5 m long in the Far Detector. Scintillation light from charged particles are captured by a wavelength shifting fiber which runs through each cell. The end of the fiber is exposed to a single pixel on an avalanche photo-diode array to record the intensity and arrival time of photon signals. The spatial and absolute response of the detector to deposited light is calibrated out using physical standard candles - this is such that a calibrated response can be derived which is a good estimate of the true deposited energy. Parallel cells are arrayed into planes, which are configured in alternating horizontal and vertical alignments to provide separate, interweaved X-Z and Y-Z views. The 14,000 ton Far Detector consists of 344,064 total channels arranged into 896 planes each 384 cells wide which are subject to external showers of cosmic neutrino and cosmic ray events.⁶ Information from the two views can be merged to allow 3D event reconstruction as depicted in Figure 5.

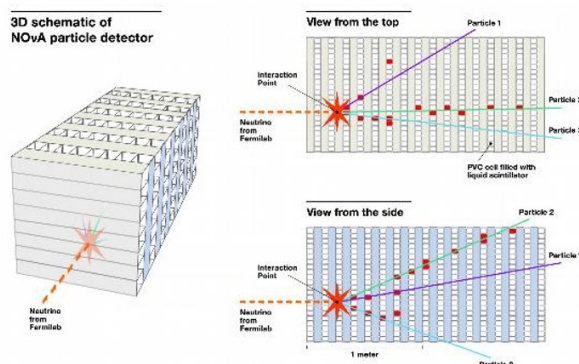


FIG. 5. The two figures on the right show the views through the top and side of the three-dimensional figure on the left. They show the hits produced as charged particles pass through and deposit energy in the scintillator-filled cells. Illustration courtesy of Fermilab.

III. PROJECT GOALS

A problem in experimental high-energy particle physics (HEP) is the correct categorization of the particle interactions recorded in the detectors as signal and background. This characterization has been done by reconstructing high-level components such as clusters, tracks,

showers, jets, and rings associated with particle interactions recorded by the detector and summarizing the energies, directions, and shapes of these objects with a handful of quantities as described above. Utilizing computer vision has made great advances by moving away from using the specifically constructed features to the extraction of features using a machine learning algorithm known as a convolutional neural network (CNN).⁷ CNNs are well suited to a broad class of detectors used in HEP and particularly in high energy neutrino physics such as in the NOvA experiment.

The near detector of the NOvA experiment is underground while the far detector is located above ground. The far detector while above ground is subject to approximately 11 billion cosmic rays per day. This presents a major issue since it is difficult to process the number of cosmic events. Approximately 10^7 of these events need to be rejected in order to properly process the pixel maps.

The goal of the project was to use an algorithm, CVN (Convolutional Visual Network), to construct a cosmic rejection network that would be capable of identifying neutrino interactions based on their topology and to reconstruct those events accurately.

IV. DEEP LEARNING AND CONVOLUTIONAL NEURAL NETWORKS

The multilayer perceptron (MLP), or traditional neural network, is a machine learning algorithm with wide and versatile use in high energy physics. The structure of an MLP consists of an input layer, one or more hidden layers, and an output layer. The goal of an MLP is to approximate a function f in a manner:

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m \quad (3)$$

where n corresponds to the dimensionality of the input \vec{x} and m is the dimensionality of output \vec{f} . All layers in traditional MLPs are fully connected; the output of each node is the weighted sum of the outputs of all nodes in the previous layer plus a bias term, operated on by a non-linear function. Traditionally, the preferred non-linearity is a sigmoid function such as hyperbolic tangent or the logistic function.⁸ An MLP with a single hidden layer, under certain assumptions, can be shown to approximate any function to arbitrary precision given a sufficient number of hidden nodes. The weights and biases used in an MLP are typically determined using a method known

⁶ A Convolutional Neural Network Neutrino Event Classifier, Aug. 12, 2016, <https://arxiv.org/pdf/1604.01444.pdf>

⁷ Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard et al., Backpropagation applied to handwritten zip code recognition, *Neural Comput.* 1 (Dec., 1989) 541551.

⁸ R. Reed and R. Marks, *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. A Bradford book. MIT Press, 1999

as supervised learning. During supervised learning, the MLP is presented examples where both \vec{x} and the corresponding output, \vec{f} , referred to as the ground truth, are known. The loss, a measure of the error between the output of the MLP and the ground truth is computed, and its gradient as a function of the weights and biases is calculated using the back-propagation algorithm. The loss is then minimized by altering the weights and biases using the stochastic gradient descent method. This procedure is repeated until the errors are reduced to an acceptable level.

The MLP is a powerful technique, but there are a number of issues with the technique. Primarily, it scales poorly to a large number of raw inputs. Historically, most of the work in developing an MLP for a particular task is devoted to extracting features from the raw data that could be used as optimal inputs. In HEP, this is essentially the process of reconstruction; however, developing optimal, robust reconstruction routines is difficult and time consuming using an MLP.

Second, although a single hidden layer can approximate most functions to an arbitrary precision, the number of nodes necessary in that hidden layer may approach infinity. Networks with multiple hidden layers can often reach the required accuracy with fewer nodes than the equivalent single layer network, and multilayer networks can be extremely difficult to train. This is partially due to the fact that as the input to the sigmoid approaches $\pm\infty$, the gradient approaches zero. The updates to the weights and biases applied using the stochastic gradient descent method have a term proportional to the gradient, so this situation can actually slow down or halt learning.

Third, the large number of free parameters in a large network runs the risk of possibly over-training - a case in which the network learns to reproduce the training sample too well and fails to generalize to inputs it has not seen before.

The focus of this project was on using a convolution neural network, a highly successful technique in the field of computer vision identification tasks. The technique was inspired by studies of the visual cortex system of animals. It was found that the visual cortex contains simple cells, which are sensitive to edge-like features within small regions of the retina, and complex cells, which are receptive to collections of simple cells and are sensitive to position independent edge-like features.⁹

These structures can be modeled by performing discrete convolutions to extract simple features across the visual field. Convolutional neural networks mimic this structure using a series of convolutional layers that extract a set of features from the input image and pooling layers that perform dimensionality reduction and add translational invariance. The data passed from layer to

layer in a CNN has a three dimensional structure - height, width, and channel number. In this case, only one channel is used. Height and width refer to the dimensions of the input image, and channel number is defined in analogy with the RGB channels of color images. For an $n \times m$ convolutional layer, the input data is transformed according to:

$$(f \cdot g)_{p,q,r} = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^c f_{i,j,k,r} g_{p+i,q+j,k} \quad (3)$$

where $(f \cdot g)_{p,q,r}$ refers to the (p, q) pixel of the r channel of the transformed image, n and m are the height and width of the convolutional kernel, c is the number of channels of the input image, f is a filter, and g is an array corresponding to pixel intensities of the input image. The filter f is a four dimensional tensor where i and j index the height and width of the filter, k indexes the input channel, and r indexes the output channel, and it is trained to identify features within the image. For a fixed k and r , the filter, f , can be thought of as an $n \times m$ convolutional kernel. After applying a separate convolutional kernel to each channel and performing a weighted sum across channel, the resulting output image is known as a feature map. The range of the r dimension determines the number of c stacks of $n \times m$ convolutional kernels that are trained. Each of these stacks of kernels produces a feature map which are stored in the channel dimension of the layer output. Finally, each output pixel is operated on by a non-linear function.¹⁰

The network utilized in this paper is inspired by the GoogLeNet architecture, which excels at ImageNet image classification task¹¹. The core of GoogLeNets power comes from its use of the network-in-network (NIN) approach to augment the learning capacity of convolutional layers while also reducing dimensionality. In NIN, the main network is composed of repeating sub-networks, where each sub-network resembles a complete convolutional CNN with convolution layers at a variety of scales to capture complex behavior. To avoid exponentially increasing the number of feature maps, NINs use a convolutional layer applying 1×1 convolutional kernels. This performs a weighted sum over feature maps to down-sample into a smaller number of maps. The sub-network in the GoogLeNet architecture, called the inception module, is shown in Figure 6.

Each branch of the inception module applies filters which extract features of various scales. The GoogLeNet

⁹ D. Hubel and T. Wiesel, Receptive fields and functional architecture of monkey striate cortex, *Journal of Physiology* 195 (1968) 215243.

¹⁰ A Convolutional Neural Network Neutrino Event Classifier, Aug. 2016, arXiv:1604.01444v3

¹¹ O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma et al., ImageNet Large Scale Visual Recognition Challenge, *International Journal of Computer Vision (IJCV)* 115 (2015) 211252

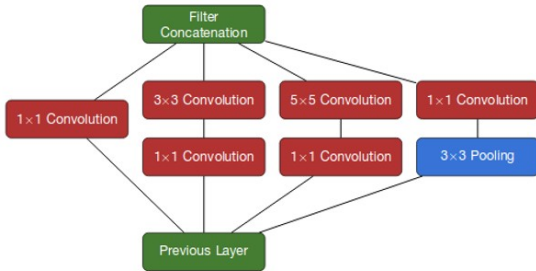


FIG. 6. . Diagram of the inception module Like any single layer, the inception module takes the set of feature maps produced by the previous layer as input. It then distributes those feature maps to branches, each with filters at different scales. NIN architecture is implemented with 1×1 convolutions to down-sample into a smaller number of maps, maintaining the dimensionality of the input maps. Separate branches perform 33 and 55 convolution, as well as 3×3 overlapping pooling.

architecture also makes use of the technique local response normalization (LRN) in which the response of a given cell in a kernel map is normalized relative to the activity of adjacent kernel maps. This creates competition for large valued features between outputs computed by different kernels which helps the network avoid local minima and to converge to a more optimal set of weights.

The neutrino flavor state can be determined in charged-current (CC) interactions which leave a charged lepton in the final state; an electron in the case of ν_e , a muon in the case of ν_μ , or a tau in the case of ν_τ . Neutral-current (NC) interactions bear no signature of the flavor of the interacting neutrino and are thus a background for the charged-current analyses, but may be signal events in other searches. To support these analyses, the CVN identifier to characterize candidate neutrino events into one of the following interaction types:

* ν_μ charged current- A muon plus a hadronic component. One of the main topological features of these events is the long, low dE/dx track corresponding to the track of a minimally ionizing muon.

* ν_e charged current- An electron plus a hadronic component. The electron topology is typically a wide shower, rather than a track, whose dimensions are related to the radiation length of the detector material.

* ν_τ charged current - A tau plus a hadronic component. The tau is extremely short lived and not visible in the NOvA detector but decays immediately with varying final state probabilities that may produce pions, electrons, muons, and neutrinos.

* ν NC- The outgoing lepton in these interactions is a neutrino, which will travel onward undetected. Thus, only the hadronic component of these events is visible, making their flavor impossible to identified. *Cosmic events - (Usually) Long muon tracks entering tops or sides

While it is useful to think about each category as a

particular iconic topology as described above, misidentification can still occur. For example, particular NC interactions can be mistaken for CC interactions when they produce pions which look like leptonic activity. A charged pion track can appear quite similar to a muon track, with the exception of a spike in energy deposition at the end of the track. As another example, a neutral pion will rapidly decay to produce a pair of photons which themselves produce electromagnetic showers, which are difficult to distinguish from showers produced by an electron. By constructing an identification algorithm like CVN, which views the entire event topology, the hope is to minimize these misidentification failure modes but they still will remain a challenge.

V. TRAINING

The CVN was implemented and was developed using the Caffe framework. Caffe, an open framework for deep learning applications, is highly modular and makes accelerated training on graphics processing units straightforward. Common layer types are pre-implemented in Caffe and can be arranged into new architectures by specifying the desired layers and their connections in a configuration file. Caffe is packaged with a configuration file implementing the GoogLeNet architecture. GEANT4 simulation tuned to data using a package called PPFX was used to simulate the flux in the NuMI beamline. Before oscillations, the NuMI beam is composed mostly of ν_μ with 2.1 percent intrinsic contamination via ν_e . It should be noted, however, that the training samples did not have the same composition as the actual beam. Neutrino-nucleus interactions were simulated using the GENIE¹² while GEANT4¹³ was used to propagate products of the neutrino interactions through a detailed model of the NOvA detectors. Custom NOvA simulation software converted energy depositions into simulated electronic signals which correspond to the detector output.

Pixel maps were created to model full detector events sliced in time to a spill window width of approximately $12 \mu\text{s}$. Any detector activity with less than 10 hits associated with a neutrino event were discarded as it would not be enough information generally to provide a positive ID of the interaction. Distinct categories were created for each GENIE CC interaction mode: (ν_μ - charged current, ν_e - charged current, ν_τ - charged current), a category used for all GENIE NC interactions to give the distribution of events, and a category dedicated for all cosmic events. An example of a pixel map can be seen in Figure 7.

¹² C. Andreopoulos et al., The GENIE neutrino monte carlo generator, Nucl. Instrum. Meth. A614 (2010) 87104

¹³ S. Agostinelli et al., Geant4 - a simulation toolkit, Nucl. Instrum. Meth. A506(2003) 250303

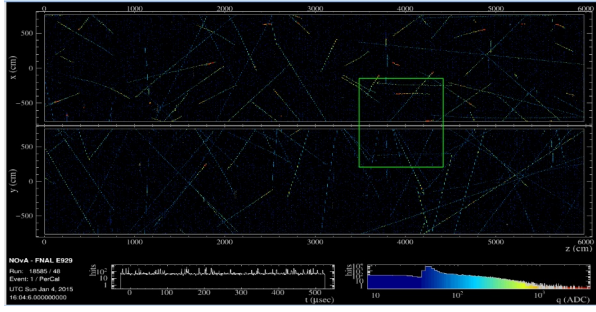


FIG. 7. . An example of a constructed pixel map which simulates the detection hits of various events including charged-current neutrino events, neutral-current neutrino events, and cosmic events. The goal of the project is to construct a network via machine learning capable of detecting and rejecting cosmic events.

TABLE I. Data composition of the training samples.

Chunk	ν_μ CC	ν_e CC	ν_τ CC	ν NC	cosmics CC	total
chunk-05	60397	49321	15921	65904	272165	463708
chunk-06	60288	52428	17651	73230	257195	460792

Large levelDB's were created which needed to be chunked. There were 250 files per LevelDB which corresponded to approximately 470000 pixel maps (over 12000 files available). In the process of training, 80 percent of the samples were used for training while 20 percent were used for testing. In this case, 376,000 were used for training while 94,000 were used for testing. The training sample for the project consisted of two separate chunks of levelDB's consisting of simulated events that were dubbed "chunk-05" and "chunk-06" respectively. This can be seen in Table 1. For "chunk-05", the composition consisted of 60397 ν_μ events, 49321 ν_e events, 15921 ν_τ events, 65904 ν neutral current events, and 272165 cosmic events. For "chunk-06", the composition consisted of 60288 ν_μ events, 52428 ν_e events, 17651 ν_τ events, 73230 ν neutral current events, and 257195 cosmic events. Validation trees were produced with a subset of 250,000 of the events. I made use of a training system in which 16 images were evaluated per iteration.

VI. RESULTS AND ANALYSIS

In order to determine how well the system was being trained to identify the difference between cosmic events and other neutrino events, an analytical graphic known as a confusion matrix was produced. In my confusion matrix (also known as an error matrix), each row of the matrix represents the instances in one of the five classes identified as:

- * ν_μ charged current- A muon plus a hadronic component. One of the main topological features of these events is the long, low dE/dx track corresponding to the track of a minimally ionizing muon.

- * ν_e charged current- An electron plus a hadronic component. The electron topology is typically a wide shower, rather than a track, whose dimensions are related to the radiation length of the detector material.

- * ν_τ charged current - A tau plus a hadronic component. The tau is extremely short lived and not visible in the NOvA detector but decays immediately with varying final state probabilities that may produce pions, electrons, muons, and neutrinos.

- * ν NC- The outgoing lepton in these interactions is a neutrino, which will travel onward undetected. Thus, only the hadronic component of these events is visible, making their flavor impossible to identified.

- *Cosmic events - (Usually) Long muon tracks entering tops or sides

while each column represents the instances in an actual class (or vice versa). As depicted in Figure 8, the network did an excellent job of identifying cosmic events correctly but it also identified the vast majority of event classes as cosmic events as well. This network utilized a 34:1 ratio of cosmic events to neutrino events. This represented a short-fall of the network, and thus it needed to be trained in order to identify and differentiate accurately cosmics from other events.

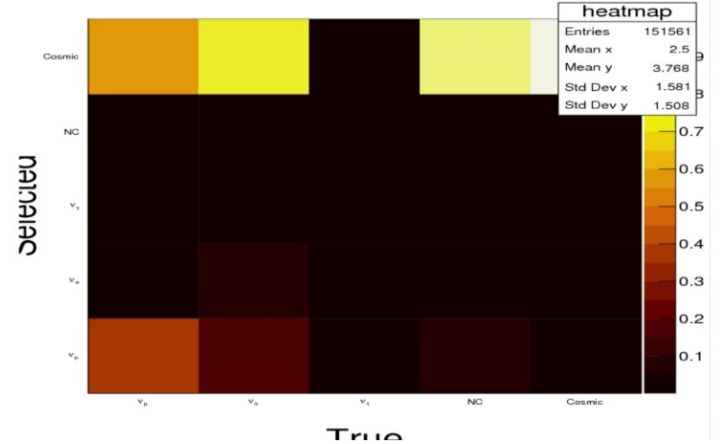


FIG. 8. . The confusion matrix of the network prior to the first training. It should be noted that this test network was severely overtrained to identify cosmic events - it was accurate in terms of identifying cosmic events but it's major pitfall was that it also identified other neutrino events as cosmic events as well!

Figure 9 depicts the accuracy versus iteration plot of training the network. The training was conducted over approximately 90,000 iterations to insure that the network was not over-trained. As can be seen, the test accuracy during training climbs significantly to an accuracy of approximately 0.76 on a scale between 0 and 1 while the training loss dropped to approximately 0.4 on a scale between 0 and 1.

In turn, this yielded the results of the confusion matrix depicted in Figure 10. The network still retained its capacity to identify cosmic events correctly and it became

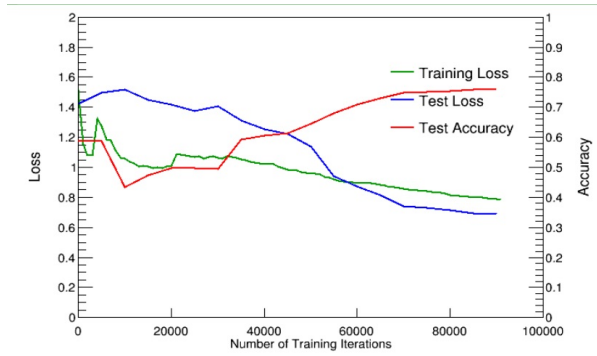


FIG. 9. The training was conducted over approximately 90,000 iterations to insure that the network was not over-trained. As can be seen, the test accuracy during training climbs significantly to an accuracy of approximately 0.76 on a scale between 0 and 1 while the training loss dropped to approximately 0.4 on a scale between 0 and 1.

progressively better at identifying the difference between neutrino and cosmic events. There was an increased accuracy in identifying muon neutrino events more accurately.

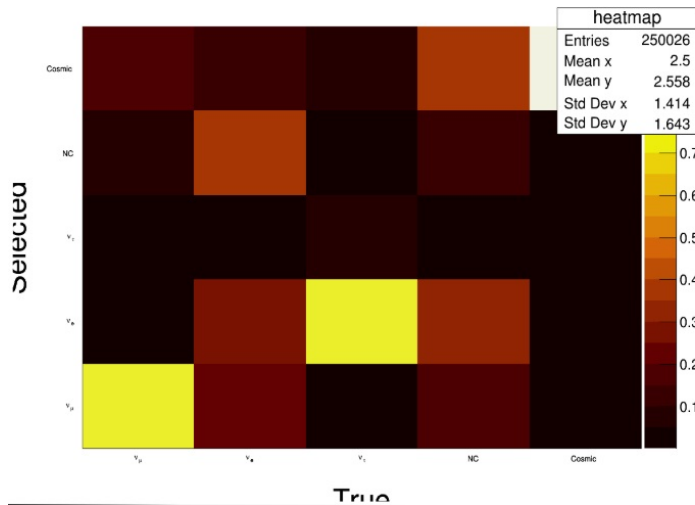


FIG. 10. . The confusion matrix yielded after training the network the first time. It was considered progressively better than the initial test matrix since it did not identify all of the events as cosmic events; of note is the increasing accuracy in identifying muon neutrino events.

Particle identification (PID) plots were generated and are used to tune selection cuts and calculate the performance of the network. The results are shown in the figures below.

Below is the PID plot for identifying neutral current events. A substantial amount of the tau neutrino events were identified as neutral current events demonstrating that the network needed to be trained to be more accurate.

Figure 15 depicts the accuracy versus iteration plot of

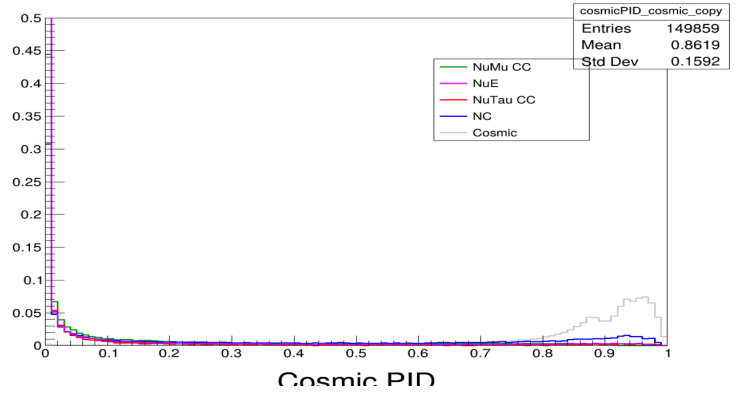


FIG. 11. PID plot for identification of cosmic events. After the first iteration of training, the network was able to accurately identify and sort cosmic events as noted by the right justified region of the histogram plot.

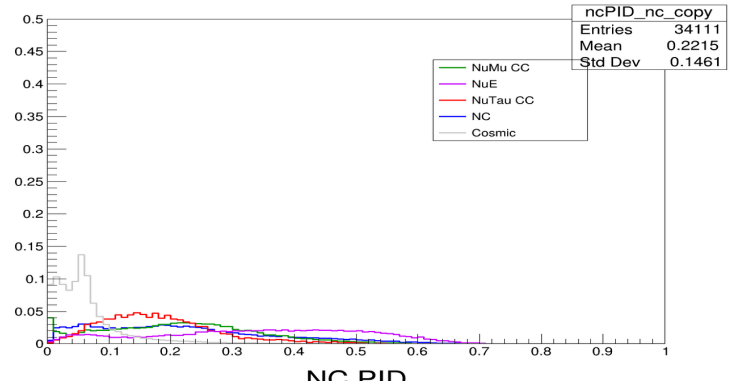


FIG. 12. PID plot for identification of neutral current events. After the first iteration of training, the network identified quite a few of the tau neutrino charged current events as neutral current events.

training the network a second time. The training was conducted over approximately 180,000 iterations which contributed to the network being over-trained. As can be seen, the test accuracy during training climbs significantly to an accuracy of approximately while the training loss actually began to climb as well. The test loss was greater than the training loss and increased as well. This exhibition of behavior was not ideal for training the network.

Shown below in Figure 16 is the confusion matrix for the network after the second training. It should be noted that this network while accurately identifying cosmic events identified a variety of events as neutral current events errantly.

Shown below in Figure 17 is the cosmic PID plot for the second training on the network. As evidenced by the peak on the histogram on the right side of the graph, the network in terms of identifying cosmic events has been tuned.

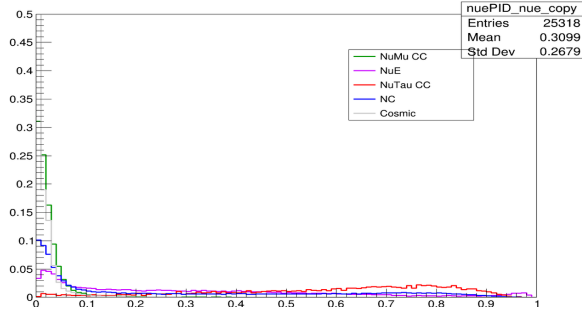


FIG. 13. . The confusion matrix yielded after training the network the first time. It was considered progressively better than the initial test matrix since it did not identify all of the events as cosmic events; of note is the increasing accuracy in identifying muon neutrino events.

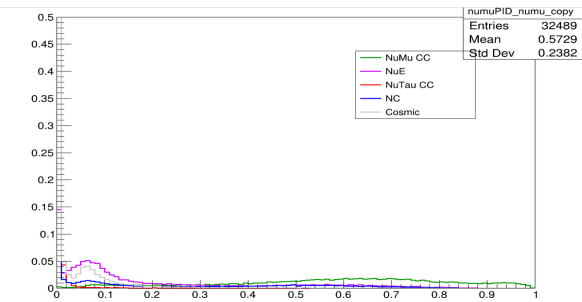


FIG. 14. The PID plot yielded after training the network the first time for muon neutrino events. Considerably, the network became progressively better at identifying muon neutrino events but still identified some of the electron neutrino events as such.

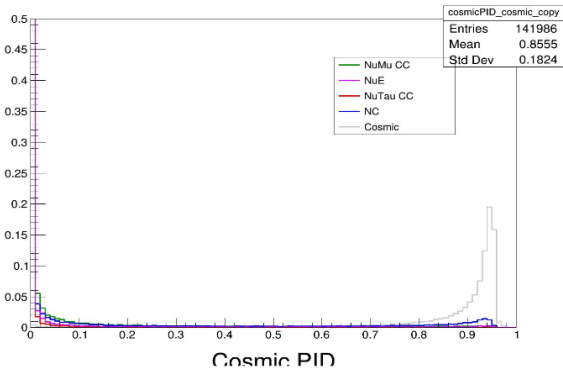


FIG. 17. As evidenced by the peak on the histogram on the right side of the graph, the network in terms of identifying cosmic events has been tuned. However, due to over-training, the network errantly identifies a wide array of neutrino events as neutral current events. The PID plot demonstrates the 83 percent accuracy in identifying and possibly rejecting cosmic events.

Shown below in Figure 18 is the cosmic PID plot for the second training on the network. As evidenced by the peak on the histogram on the right side of the graph, the

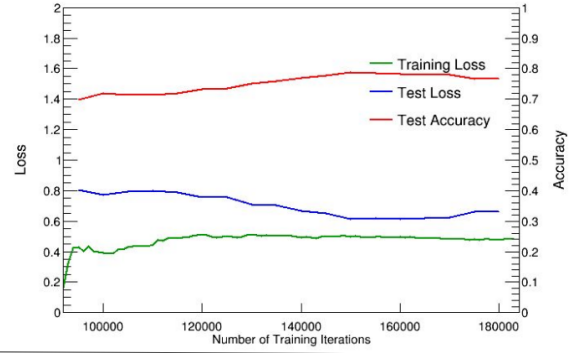


FIG. 15. The accuracy versus iteration plot demonstrating the second round of training on the network. The network experienced some over-training which leaves it susceptible to generalization of the various labels used as identifiers in the confusion matrix.

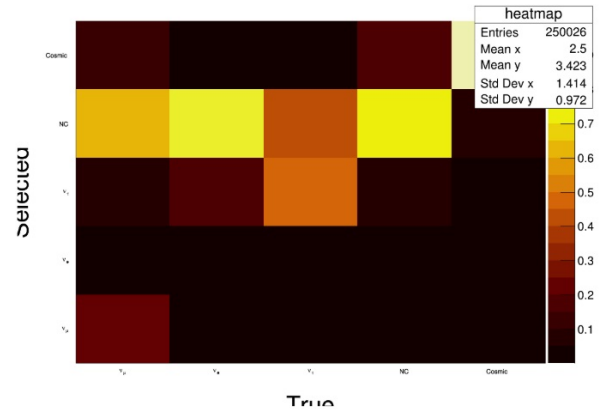


FIG. 16. The confusion matrix for the network after the second training. It should be noted that this network while accurately identifying cosmic events made the error of identifying a variety of events as neutral current events errantly.

network in terms of identifying cosmic events has been tuned.

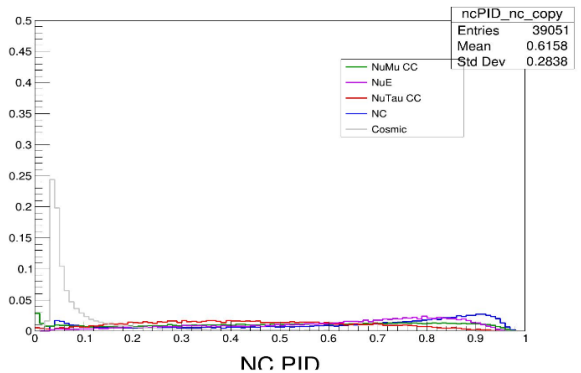


FIG. 18. The PID plot for neutral current events. Of note is the 89 percent accuracy in rejecting cosmic events.

Shown below in Figure 19 is the cosmic PID plot for

the second training on the network.

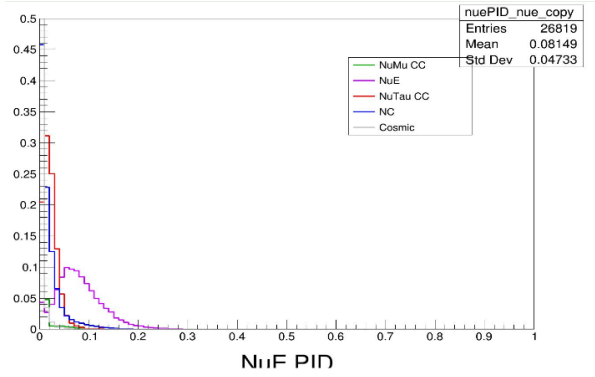


FIG. 19. The PID plot for identification of electron neutrino events.

Shown below in Figure 20 is the ν_μ PID plot for the second training of the network.

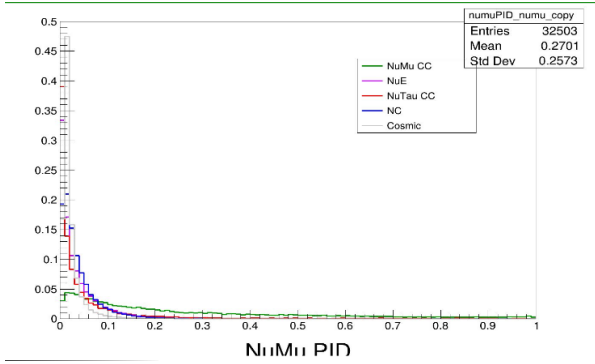


FIG. 20. The PID plot for muon neutrino events.

In order to determine the efficiency of the cosmic rejection network, it was necessary to plot the "cut value" based on a bin range between 0 and 100 versus the efficiency in sorting each event. For the network to be considered successful and ready for implementation, it was required that greater than 99 percent of the cosmic events needed to be rejected. The efficiency of ν_e and ν_μ neutrino events has to be greater than 99 percent to be considered competitive and accurate. Pictured below are the graphs corresponding to these cut versus efficiency measurements. For the region corresponding to a cut greater than 0.9 for the cosmic plot which is needed to reject the cosmic events, the accuracy is approximately 60 percent. This accuracy does not meet the criteria for successfully rejecting cosmic events. However, with an accuracy greater than 50 percent, it does show promise as an implementable algorithm.

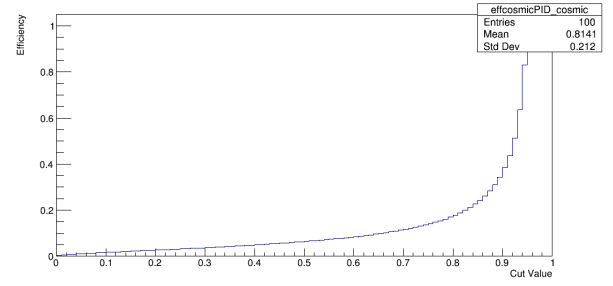


FIG. 21. Cut versus efficiency plot for cosmic events. For the region corresponding to a cut greater than 0.9 for the cosmic plot which is needed to reject the cosmic events, the accuracy is approximately 60 percent.

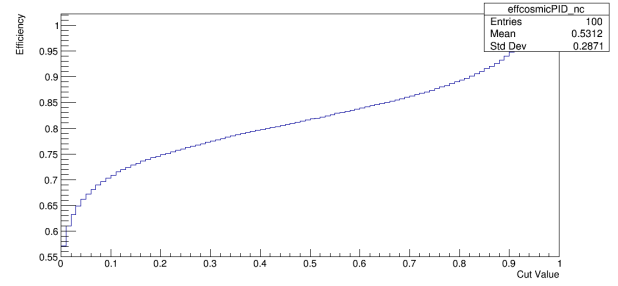


FIG. 22. Cut versus efficiency plot for neutral current events.

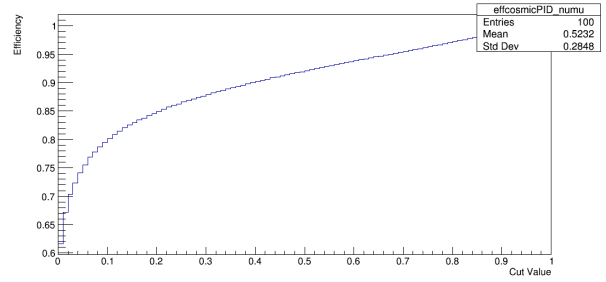


FIG. 23. Cut versus efficiency plot for electron neutrinos.

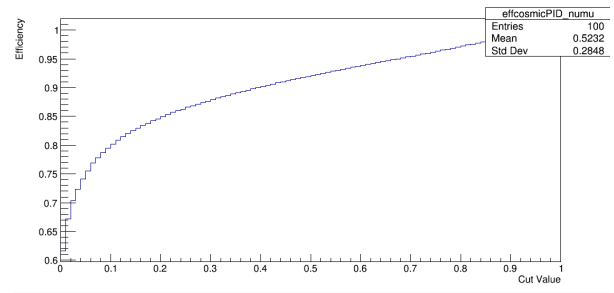


FIG. 24. Cut versus efficiency plot for muon neutrinos.

VII. CONCLUSION AND FUTURE WORK

In conclusion, there needs to be continued tuning and training of the network in order to make it a viable and integrable tool for the purpose of rejecting cosmic events for reconstruction of individual neutrino events. The network demonstrates promise as an analytical tool for future use.

In the future, plans include utilizing multi-access DB's and moving to machine learning packages such as TensorFlow for future use.

VIII. ACKNOWLEDGEMENTS

I would like to express my sincere gratitude towards Fermi National Accelerator Laboratory for offering this research opportunity through funding and opportunities made possible by the National GEM Consortium. I would also like to express my gratitude towards Dr. Alex Himmel and Dr. Evan Niner for their continued support, guidance, and tutelage that made this project possible, and for the opportunity to continue to contribute to research for the NOvA experiment. I would like to thank my fellow colleague, Adam Moren, for his continued support and assistance with the project. I would also like to thank the U.S. Department of Energy for its continued support and funding of STEM-based pursuits. A monumental thanks is in order to the GEM Consortium for supporting me as a fellow and for funding the research and contributions developed on this project. Finally, to my friends, family, colleagues, mentors, and those who played a role in being the chief support of my dreams, goals, and ambitions while also providing guidance, mentorship, and advice, I would like to express a sincere amount of gratitude for your presence and actions. "It takes a village to raise a child." How much more so does it take that village to support one of the budding members of its community!

IX. REFERENCES

R., H., M., P., S., V. (2016, August 12). A Convolutional Neural Network Neutrino Event Classifier. Retrieved

from <https://arxiv.org/abs/1604.01444>

A Convolutional Neural Network Neutrino Event Classifier. (2016, August 12). Retrieved from <https://arxiv.org/pdf/1604.01444.pdf> R., H., M., P., S., V. (2016, August 12).

A Convolutional Neural Network Neutrino Event Classifier. Retrieved from https://arxiv.org/abs/1604.01444v3?utm_campaign=Revuenewsletterutm_medium=Newsletterutm_source=revue

A., A., R., W., N., Y., . . . Dell'Acqua. (2002, August 22). GEANT4: A Simulation toolkit. Retrieved from <http://inspirehep.net/record/593382?ln=en>

Hammer, B. (2001). Neural Smithing Supervised Learning in Feedforward Artificial Neural Networks. *Pattern Analysis and Applications*, 4(1), 73-74. doi:10.1007/s100440170029

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1), 215-243. doi:10.1113/jphysiol.1968.sp008455

R., O., D., J., H., K., . . . L. (2015, January 30). ImageNet Large Scale Visual Recognition Challenge. Retrieved from <https://arxiv.org/abs/1409.0575>

Lecun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4), 541-551. doi:10.1162/neco.1989.1.4.541

Neutrino 2014: The 24th Interactional Conference on Neutrino Physics and Astrophysics. (2014, June). Retrieved from <http://neutrino2014.bu.edu/program/>

M., G., W., R., R., B., . . . Valencia. (2016, July 05). Neutrino Flux Predictions for the NuMI Beam. Retrieved from <http://inspirehep.net/record/1473668>

Research goals — NOvA. (n.d.). Retrieved August 5, 2018, from <http://novaexperiment.fnal.gov/research-goals/>

A., B., C., D., D., G., . . . Bell. (2009, November 18). The GENIE Neutrino Monte Carlo Generator. Retrieved from <https://arxiv.org/abs/0905.2517>