# A Comparative Review of Neural Networks for Neutrino Detection

Shruti Rao
Vrije Universiteit, Amsterdam & University of Amsterdam
s2.rao@student.vu.nl

## ABSTRACT

Particle physics involves careful examination of sub-atomic particles and their interactions with each other and their surrounding elements. The main challenge across particle physics remains the separation of background noise from the true signals. Physics controls and algorithms that were being used for examination so far have been successful to a certain extent. With the rapidly increasing volumes of data generated, these algorithms are unable to keep up or are hard to improve such that new findings can be made possible. These concerns paved way for neural networks to take over the traditional physics algorithms, due to their ability to handle and learn complex, non-linear relationships. This transition to neural networks is especially being considered in neutrino studies.

There is a lack of systematic information on what neural networks have been used in neutrino research and to what degree of success. This study examines relevant work done with neural networks in the field of particle physics. The study has a two-fold interest. It wishes to aid physicists who may be lacking knowledge in artificial intelligence to use this study as a guide on getting started with neural networks in their field. It also wishes to serve as a reference point for computer scientists who are looking to develop new learning techniques or architectures that can specifically cater to this niche yet significant branch of physics.

The study identified notable reliance on convolutional neural networks (CNNs) in the field. It is evident from the studies that CNNs can be used as a valid starting architecture as it has already provided successful results for existing experiments. The study can also recommend Graphical Neural Networks (GNNs) for geometrically irregular data based on it's use at a detector. The study noted challenges in objectively comparing results across studies due to differing metrics and recommends a standardisation be put in place. Overall, despite the relatively few case studies using neural networks in particle physics, it is evident that these networks can be a part of future physics research.

## KEYWORDS

systematic literature review, deep neural networks, convolutional neural networks, particle physics, HEP, neutrino detection, jet analysis

## 1 INTRODUCTION

Physics through the course of time has accounted for many fundamental properties of the universe. Yet, several questions regarding the elementary constituents of matter still remain unanswered. For instance, it is well known that when neutron stars collide, they produce supermassive stars or black holes [13]. However, there is not much information on what the cores of such stars or black holes comprise of. What is however known, is that all of these events have one particle in common - a neutrino [13]. Neutrinos are elusive, weakly interacting particles that were discovered first by Pauli in the 1930s [16]. Majority of the universe is made of dark matter - what scientists believe to be the key to understanding the origins of universe [29]. Neutrinos are the only known particles from dark matter [38]. Understanding neutrinos has become increasingly significant for researchers. Specifically, experiments are being conducted to understand the mass of neutrinos, the reason for their oscillation, their ability to change forms and the role it plays in the birth and continuum of the universe [27].

With simultaneous advancement in hardware and computing power, the ability to detect and understand neutrinos on Earth has drastically increased. Particle physics has taken upon the role of understanding neutrinos and the laws that govern it [2, 3, 7, 8, 16, 38]. Large particle accelerators are atypical for particle physics experiments whereby protons and anti-protons are collided at high speeds to try to recreate exotic particles. These exotic sub-atomic particles are such that they can hardly ever be observed directly. Instead, detectors look to capture evidence of interactions these particles have with each other or with their surroundings [16]. Through this process, petabytes and even exabytes of data are collected in real time and analysed for signs of tracks, rings, jets and showers that are associated with such particle interactions [35]. Such experiments so far have made use of physics algorithms and those have worked well in detecting particles to a certain degree. However, these techniques fail when it comes to identifying new particles or studying previously unknown behaviour that has not been defined by the algorithm parameters [26]. These algorithms are also unable to keep up with large volumes of rapidly changing data. Thus, reliance on physics algorithms have led research to a standstill and limited the potential for new discoveries [26].

Meanwhile, neural networks have faced several cycles of hype over the past decade or so. Early attempts at incorporating neural networks were often unsuccessful due to limited understanding, large computational hours, hardware limitations, and lack of powerful architectures [6]. Early applications that were developed were highly sensitive to errors and data quality [6]. They were unable to cope with changing data and varied, complicated data types [6, 26]. At the time, combining expensive particle detectors with such fickle systems were viewed as an inconvenience rather than an advantage. Moving onto recent times, significant advances in computing power led to better storage of data, faster execution times and improved error handling. These directly contributed to allowing real-time processing of data. Storage of large datasets were now made possible, directly affecting the ability to train networks with larger datasets. Additionally, general theoretical research was mitigated to address the concept of neural networks being a black-box. Advances in computational theory led to development of powerful learning algorithms, optimisation techniques and robust architectures [17].

These factors combined led to new interest in application of neural nets to complicated problems, including problems in particle physics [26].

## 1.1 Background on Neutrinos

Neutrinos are fundamental particles of the universe and considered exciting because of how different they are from other particles. Unlike other elementary particles, they carry no charge and are extremely small in mass, close to zero [16]. They also do not interact via the strong forces and electromagnetic forces. Rather, they interact only via the weak subatomic force - a subatomic force that causes radioactive decay of atoms. Thus, they travel through matter undetected [16]. They come in three flavours or types called *electron neutrinos*, *tauon neutrinos* and *muon neutrinos* and it was recently discovered that neutrinos can change types and masses [16].

Most of the neutrinos present are known to have existed soon after the formation of the universe. On Earth, neutrinos are produced by nuclear reactors, natural radioactive changes in the atmosphere and particle accelerators. The Sun produces neutrinos via nuclear fission that occurs in its core. They are also generated from the births, deaths and collisions of stars and supernovae explosions [38].

A trillion neutrinos approximately pass through the Earth every second yet, only one neutrino gets to react with matter on Earth, once every day [4]. For a neutrino to react, it has to hit an atom at it's core. When it does so, it results in *weak boson* particles. However, detecting them remains extremely hard for two reasons. First, when neutrinos interact to give out resulting particles (*weak bosons*), they only last for one ten thousandth of a trillionth of a trillionth of a second [4]. Second, these particles travel a distance of less than one-one-thousand of a size of a proton. [4] These two factors combined make it extremely challenging for a neutrino to react with an atom and thus get detected.

There are a few ways of detecting neutrinos with under-water detectors being most popular. In water, neutrino particles travel undisturbed and may travel faster than light. They may react with some particles in the water and create a charged *lepton* that produces a light known as *Cherenkov's light*. These flashes of light are detected by photomultiplier tubes that can infer direction, energy and flavour of the neutrino [4]. Detectors additionally need to cover a significant surface area so that more than one neutrino can be detected per day. No matter the kind of detector, they all need to be placed such that background noise from cosmic activity and other terrestrial noise is minimised. Several experiments have been set up that attempt to do just this. *Super Kamiokande* is a water based detector that uses *Cherenkov light* to detect neutrinos. It was successfully able to detect neutrinos from a Supernova, which led to renewed interest in the field [44]. *IceCube* is another experiment located in the South Pole that uses a cubic kilometre of ice embedded with photomultiplier tubes to detect neutrino events [8]. *MiniBooNE* detector uses pure mineral oil that allows low energy muons and protons, invisible in water, to be detected [5]. The *KM3NeT* is another effort currently being built under the Mediterranean sea, focused on detecting deep space neutrinos [7].

Neutrinos are the most abundant particles in the universe but hardly much is known about them. Studying their origins can help resolve many mysteries of the universe. Since they travel through space practically unaffected, physicists believe that neutrinos can help learn about the origins of universe.

## 1.2 History of Neural Networks in Particle Physics

Neural networks (NNs) were first acknowledged in physics around 1988, in the field of particle physics [23]. Particle physics comprises the study of the fundamental building blocks of nature - quantum physics, irreducible, elementary particles and big bang [2]. Particle physics largely involves low level pattern recognition and physics process determination. According to Denby (1999), low level pattern recognition includes finding tracks made by particles and process determination encompasses obtaining properties such as angular momentum, spatial topology and energy emissions of particles. Studies of such processes require work to be done either in real time or offline. Denby (1999) described particle physics processes to be characterised by larger magnitudes of background noise with small, rarer occurrences of real events at any given point in time. Therefore, data analysis are of two kinds - real time or offline research. Applications that make use of real-time triggers, attempt to filter out most background noise and look out for particle events in real time. Offline reconstruction requires using massive compute power to build such events and employ efficient algorithms to parse through the noise [23].

Denby (1999) found neural networks that have been used in particle physics have been used in both real time and offline applications. Overall, neural networks have had challenges being recognised as a statistical tool within the community of particle research. The main challenge in particle physics lies in the fact that such experiments often have to deal with new and unknown phenomenon. Neural networks in such instances have to be developed based on unknown, and guessed parameters. Models trained on such parameters then further reflect these unknowns and inaccuracies. The ease of use of neural networks make it tempting to combine them with unknown variables to obtain results, and this remains the biggest trap in the field [23]. Given these accepted fallacies, there are a few large-scale detector experiments that have attempted to incorporate neural networks. *Fermilab* has a muon trigger built around a test beam in it's detector that applies low level pattern recognition techniques [4]. *Fermilab* also uses neural networks to analyse proton-anti-proton collisions and measures top-quarks and lepto-quarks [4]. The *Hera* accelerator has a prototype experiment that studies momentum from colliding particles [45]. The *Hera* accelerator has a secondary experiment called *ZEUS* that uses a form of feedforward network to identify deeply inelastic neutral current events [1]. The *CMS* experiment at the Large Hadron Collider (LHC) uses a neural network based trigger to identify electrons from protons [18]. The transition radiation detectors (TRD) at *CERN* use pattern recognition systems to discriminate between electrons and hadrons whereby particles are identified based on the electromagnetic radiation patterns produced [14].

## 1.3 Contribution

The goal of this study is to conduct a systematic assessment on applications of neural networks in the field of particle physics. The results of this assessment can determine if neural networks are indeed a promising solution for further research of neutrinos.

In order to achieve this goal, a systematic collection of studies were gathered and finalised with the help of additional filters. For each selected case study, the area of focus for the paper was determined. Next, the problem space was identified and all preprocessing and data preparation was noted. Special attention was given to the architecture and training decisions. Relevant metrics were noted and efforts were made to understand if the neural network provided any advantage over existing methodologies.

At present, there are no comparative reviews that have assessed the state-of-the-art for application of neural networks to neutrino research (and by extension particle physics). With growing body of research in particle physics and artificial intelligence, it is safe to predict that there may be more researchers who would wish to adapt neural networks to their own work. This study aims to assist physicists lacking in-depth AI expertise to assess the work that has already been done in their field and provide a reference point. Results from this study could also be examined by artificial intelligence experts to understand the gaps between what particle physics needs and what neural networks can deliver, and work on developing more streamlined solutions for the community.

The rest of the paper is organised as follows - study design (Section 2) for this work is described through the combination of research goal and research questions. Next, to answer these research questions, a search strategy and criteria are determined to find suitable papers. Data extraction procedure is then briefly stated and the validity of this study is highlighted. Results (Section 3) from primary studies are organised by topics of particle physics and discussed. These results are summarised (Section 4) and the study is concluded (Section 5) by stating potential implications for future work.

# 2 STUDY DESIGN

## 2.1 Research Goal

Preliminary search demonstrated insufficient applications of neutrino specific neural networks and thus, the scope of this study extends to neural networks applied across all constituents of particle physics. The results are a valid extension to neutrinos as they form a subset of particle physics [38].

This study aims to carry out a systematic review of the state-of-the-art neural networks in the field of particle physics to identify how they have been applied. The study aims to identify a set of suitable candidates for the purpose of neutrino detection and highlight the conditions under which they were deemed successful. The overview of case studies in this paper can allow researchers to decide if they can use existing methodologies, or further develop them for their work. Additionally, computer scientists can use this study to determine the shortcomings of the present algorithms and architectures and develop new ones based on the requirements in field.

## 2.2 Research Questions

The goal of this study is to examine and summarise relevant neural networks that are in use in the field of particle physics. To assist with the stated research goal, research questions were first formulated.

**RQ-I** What are the types of neural networks that have been applied in particle physics?

**RQ-I.I** What kind of analysis has being conducted using these neural networks in particle physics?

**RQ-I.II** Are there any neural networks that have been used to specifically detect neutrinos?

**RQ-II** Has application of neural networks resulted in improved metrics over previous research methodologies?

*RQ-I* is the first step of the research - to identify the various architectures that have already been put to use, and forms the basis of this study. These architectures will be examined in detail in this study to understand the conditions of their setup and if they can be replicated. *RQ-I.I* notes if any specific kinds of particle analysis are more popular and work well with neural networks. Alternately, there could also be certain analysis that could be deemed unsuitable for neural networks. *RQ-I.II* specifically understands if neutrinos are currently being researched using neural networks. *RQ-II* is the quantitative aspect of the study where the study conclusively tries to determine if neural networks have positively impacted research. Positive improvements could be in the form of improved or simplified research process at the least, with proof of lesser research hours required to achieve the same results as traditional methods. It could be in the form of faster processing of the same data or ability to process larger amounts of data. Improvements in the best case would be if new information was brought to light via neural nets that could not have been discovered otherwise, leading to breakthroughs.

## 2.3 Search Strategy and Criteria

Based on the defined research goal and research questions, a search strategy was adopted (Table 2). As part of the search strategy, an electronic search space was identified, along with a list of keywords to effectively obtain all empirical studies. The list of main keywords were identified as *physics, neural network, particles*. Based on this list, several iterations were generated to include search strings and additional keyword synonyms (Table 1).

To ensure quality research, only peer-reviewed journal publications and conference papers were chosen as part of the search criterion [41]. Thus, all articles, newsletters, books, magazines and demo papers were excluded. Since neural nets were first mentioned in 1988, the search period was extended from 1988 till present date [23].

## 2.4 Selection criteria

The study focuses particularly on the use of neural networks in the branch of particle physics. To identify studies that could directly meet this goal, an inclusion and exclusion criteria was developed [31]. This is described in the Inclusion/Exclusion Criteria (I/E) below.

**I1** Studies focused on describing neural networks in the field of particle physics. This criteria was used to exclusively examine particle and matter study that used neural networks.

**I2** Studies that involved pattern recognition, image reconstruction, event classification or physics process determination. This criterion was utilised to improve relevance as identification of neutrinos and similar particles often involve

| | |
|---|---|
| **Keywords** | physics, HEP, particle physics, neutrinos, neutrino physics, neural networks classification, deep neural networks, artificial neural networks, prediction, neural networks |
| **Search String** | physics prediction and classification, physics prediction or classification, HEP and prediction, HEP and classification, HEP and artificial neural networks, HEP and neural networks, particle physics and artificial neural networks, particle physics and neural networks, neutrinos and artificial intelligence, neutrinos and deep learning, neutrinos or hep or particle physics and deep learning or neural networks or classification |

**Table 1: Search Keywords and Search Strings Used for Identification of Relevant Studies**

| | |
|---|---|
| **Search Space** | ACM Digital Library, Google Scholar, IEEE Xplorer, ScienceDirect |
| **Publication Type** | Journals, Conferences |
| **Language** | English |
| **Publication Period** | 1988 - present |

**Table 2: Search Strategy Adopted for Identification of Primary Studies**

reconstruction of paths, energy inferences, and classification against background noise.

- **I3** Studies that provided quantitative evidence of model accuracy or performance. Studies without such metrics were dismissed due to insufficient information.
- **I4** Studies that included description of network architecture. As reporting the setup was an important part of this study, only studies that included this information were retained.
- **E1** Studies that did not use neural networks as their primary methodology but as an extension. This criteria was required as this study focuses only on use of neural networks.
- **E2** Secondary or tertiary studies such as literature reviews or surveys. This exclusion criterion was adopted in order to exclude studies which did not report the desired level of detail regarding implementation.
- **E3** Studies in the form of editorials and tutorial, short papers, and poster. These were mostly for general information and thus deemed to not provide the required level of detail.
- **E4** Studies that were not published in English language. Translation of non-English text would have made the analysis time consuming and prone to misinterpretation.
- **E5** Studies that had not been peer reviewed. In order to ensure that a certain industry standard was met, peer-reviewed papers were considered as an indicator of such quality.

| **Field** | **Description** |
|---|---|
| Identifier | [Unique ID for paper] |
| Title | [Title of primary study] |
| Author | [Author (s)] |
| Year | [Year the study was published] |
| Abstract | [Short summary of study] |
| Keywords | [List of relevant keywords] |
| Search Scope | [Specify if study is published as journal or conference paper] |
| I/E Criterion | [Check against all I/E criterion] |
| Included? | [Study included if all inclusion criterion are met and no exclusion criterion are met] |
| Theme | [Specify the area of particle physics the study is relevant for] |

**Table 3: Data Extraction Form Fields and Description of Fields**

Thirty three papers were identified using the search strategy mentioned in 2.3 and were refined to fifteen papers after examining against the Inclusion/Exclusion criteria specified under 2.4. These fifteen studies formed the primary studies for this review.

## 2.5 Data Extraction

To extract and catalogue relevant information from identified primary studies, a data extraction sheet (Table 3) was used [30]. The data extraction form was designed keeping in mind the need to collect information such that the research questions could be addressed [31]. Apart from cataloguing the studies by authors, title and abstract, all papers were cross-checked against the inclusion-exclusion (I/E) criteria. The studies were marked to be included if they met all inclusion criteria and none of the exclusion criteria. Additional notes was made to asses if the studies had been published as part of a journal or a conference. Finally, the themes of particle physics that were covered by the studies were noted. This was to ensure that applications across a variety of particles were discussed at the very least.

Figure 1 shows that majority of the finalised primary studies were all from Journals. The few papers that were submitted to conferences were also submitted and published in Journals but are indicated separately in Figure 1.

Figure 2 shows the years the primary studies were published across. Majority of the chosen papers were all fairly recent work. The earliest dated paper was from 1993 which is when neural networks were initially broached upon for interdisciplinary applications [38]. It seems that the application of these networks were forgotten until 2014 whereby it picked up popularity once again. Most work towards neural networks in particle physics occurred in 2016 but then saw a gradual decline. This could be a possible indication that interest in neural networks have tapered off in the field once more. However, the papers included extend up to 2019, indicating a relevant coverage.

It was considered important to note the areas of particle physics that were covered by the primary studies. Figure 3 helps highlight the potential topics that have already been addressed as a suitable

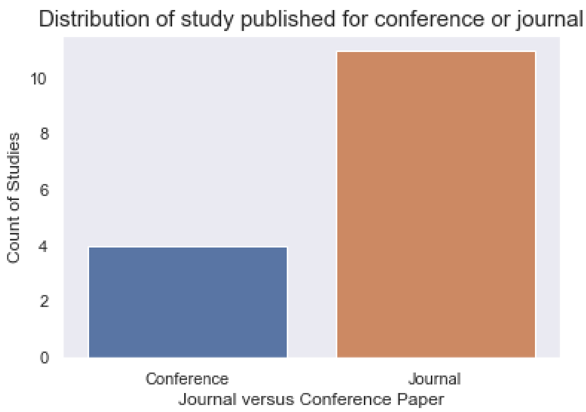


**Figure 1**

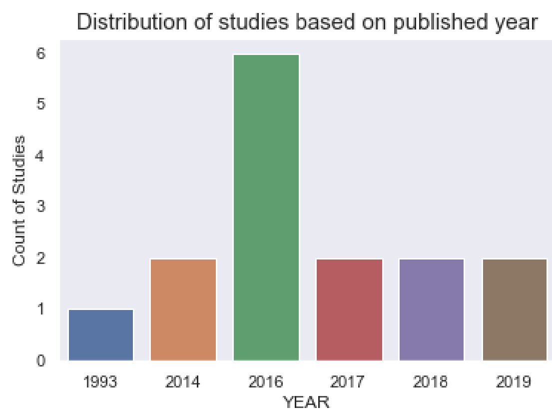


**Figure 2**

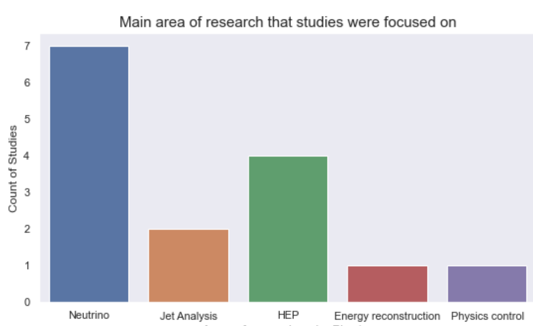


**Figure 3**

candidate for neural networks. While most topics are directly related to the particles themselves, a study seems to have applied neural networks as a control.

## 2.6 Study Validity

The ideal next step should have been to check for validity of chosen studies with the help of an expert or the publishing authors. As per Kitchenham (2007), data extraction should be performed independently by two or more researchers. Reports from participating researchers should then get compared and assessed. Uncertainties about any primary sources should be investigated as part of sensitivity analyses. Alternatively, a test-retest process could have been used where an external researcher performed an extraction from a random selection of primary studies to check for consistency. However, this step of validity remains missing in this study [31].

# 3 RESULTS

Particle identification and categorisation is core in particle physics. Common practice for characterising such particles include reconstruction of clusters, tracks, jets, rings and showers associated with particle interactions [10]. Traditional physics techniques while successful have trouble with correctness in reconstruction of such high level features. More so, the features used to characterise these events are limited by what is already known to the physicists. These factors combined limits the potential for discovering new information [10]. It was noticed that most work applied neural networks to topics of High Energy Particles (HEP), jet analysis, energy reconstruction, physics trigger mechanism(control) and neutrinos. The results from the primary study are presented here and grouped by these aforementioned themes of particle physics they explored.

**High Energy Particles (HEP)**

Baldi et al. (2014) discussed the nature of discovering particles in HEP whereby a small subspace of extremely high dimensional data has to be isolated. The key challenges in the field arise due to the fact that exotic particles are very rarely produced, exist for very short periods of time and cannot be directly observed. The hypothesis for a new particle gets tested on the subspace and the prediction gets compared against the null hypothesis. Baldi et al. (2014) state that the ratio of the sample likelihood functions for the two hypothesis are the optimal distinguishing quantity. This ratio is known as the relative likelihood. The authors found significant improvements and advances in deep learning to be a strong argument for using deep neural networks (DNN) for particle classification [12].

Baldi et al. (2014) set up benchmark cases for two kinds of particle experiments. The first benchmark classification task was defined to identify new Higgs Bosons (HIGGS). The authors obtained a set of low and high level features. They noted the superior discriminating power of the high level features. The second benchmark task aimed to distinguish supersymmetric particles (SUSY). Due to the nature of these particles, creating low and high level features were quite challenging. The authors used multivariate analysis (TMVA) to generate the baseline performance. To train their deep learning model, Baldi et al. (2014) choose 2.6 million training samples and 100,000 validation samples. They used a five-layer neural network with 300 hidden units per layer. For their parameters, they chose a learning rate of 0.05 and a weight decay coefficient of 1 x $10^{-5}$. They pre-trained their data using autoencoders. Separate classifiers were trained for each type of feature set: low level, high level and combined features, to note whether the neural network learnt the distinguishing features. This was done for both Higgs Boson and SUSY data [12].

Baldi et al. (2014) reported the area under the curve (AUC) as their primary metric as it could explain their best classifiaction model. They also calculated discovery significance - a standard metric in HEP. This metric indicates that small increases in AUC represent significant enhancement in discovery significance. The TMVA baseline classifier trained on the combination of high and low level features showed the highest AUC scores of *0.81*. In comparision, the DNN classifier resulted in a score of *0.88*. The DNN also scored higher than the *TMVA* across low and high level feature sets, with an overall *8%* improvement. A similar yet slightly less significant performance improvement was noted with the SUSY data [12].

Overall, Baldi et al. (2014) found that deep learning techniques were able to discover insight from high level features. They were able to find additional separation power, as demonstrated by the superior performance of the DNN with low-level features. The DNN demonstrated its ability to select events near the same signal values and also retain events away from the signal-dominated regions [12]. The paper was thus able to effectively demonstrate the improvement caused by DNNs over traditional *TMVA* scoring techniques. The authors did not however discuss the problems in case of multiple background noise sources. They also failed to address and provide reasoning for their hyper parameter selection.

Particle physics experiments often involve long exploratory processes of combing through high volumes of data in attempts to identify and analyse relevant signals while attempting to uncover new physics phenomenon. Racah et al (2016) discussed how relevant information could be extracted from raw data into high level representations using deep neural networks. They used data from the Daya Bay Neutrino Experiment and demonstrated convolutional deep neural networks (CNNS) as a classification filter. The Daya Bay Experiment has been trying to detect and observe anti-neutrinos produced by nearby nuclear power plants [9].

Racah et al. (2016) believed that since neural networks could express complex data, deep learning in particular would be useful for exploring high dimensional data in new light. They obtained data pertaining to charge deposits in each of the photo-multiplier tubes (PMTs) in the anti-neutrino detector. The authors attached physics derived labels to identify five types of event classes such that anti-neutrino events were separated from the rest. Since the data was in the form of 2D images, the authors first used supervised CNNs. Equal representation of the types of classes were maintained and each value in the 8x24 image was log transformed. This was done since a specific muon class had significantly higher values than the rest, skewing the overall data. The supervised CNN was initialised with convolutional layer alternated with a pooling layer and finally two fully connected layers. Altogether, the network had six layers. tanh, max and softmax were used as activation functions respectively. The network was trained on 45,000 examples using stochastic gradient descent (SGD) as an optimiser. 15,000 test samples were then trained and t-SNE was used to visualise the features. The accuracy scores across the five types of classes were consistently high (*89.1% 97.4% 99.7% 95.1 % 92.8%*) and CNN scored highest when compared with k-Nearest Neighbours (kNN) and Support Vector Machines (SVM) [36].

For unsupervised learning, Racah et al. (2016) used an autoencoder with transposed convolutional layers. Here, the encoding segments comprised of convolutional and max-pooling layers, fully connected hidden layers and finally de-convolutional layers. *Sum of squared error* was used as the loss function and the network was trained using gradient descent with a learning rate of *0.0005* and momentum coefficient of *0.9*. The network was trained on 31,700 examples and tested on 7,900 samples. t-SNE visualisations were used to interpret the effectiveness of unsupervised learning. The authors remarked on the well-defined clusters that represented the individual classes without having used any physics knowledge. The authors also compared several event images with the ones reconstructed by the autoencoders. Overall, they noted that the auto-encoder was able to filter out input noise and reconstruct the shape for different events [36].

Authors Racah et al (2016) applied CNNs to identify different classes of events including anti-neutrinos from images captured by the Daya detector. They additionally attempted unsupervised learning and noted the effectiveness of the algorithm that worked even without physics knowledge. Unsupervised learning for clustering could be a potential application as highlighted by the authors. They could further the autoencoder approach by attempting alternative filters and especially feed higher volumes of data. The authors do not discuss image resolution used but, they could test the effectiveness of unsupervised learning with variable image resolution as well. If unsupervised networks could be further strengthened, then pipelines could incorporate them as it reduces the need for complex, handcrafted features.

### Physics Control

Experiments in particle physics is predominantly accelerator based with a number of complex, non linear, interacting systems, long cycles and very small tolerance to parameter changes [38]. Traditional physics control techniques are increasingly becoming inadequate in managing such systems. Edelen et al. (2016) found neural networks to be the next generation of controls. Edelen et al. (2016) found that neural networks (NNs) were a superior tool for modelling, controlling and analysing complex, evolving systems. They describes a use case at the Fermilab Accelerator Science and Technology Facility (FAST) that they believed to be a suitable use case for application of neural network based controls. The authors identified the regulation of the resonant frequency of the electronic gun at FAST. They stated several challenges associated with the given system - specifically the cavity temperature. Traditional regulation controls failed to perform under long periods of dynamic conditions. Further the controls required manual adjustment to deal with the dynamically changing conditions, thus reducing efficiency [26].

As the problem was unconstrained and non linear, Edelen et al. (2016) used Broydon-Fletcher-Goldfarb-Shanno (BFGS) algorithm optimization to generate weights and biases. A feedforward architecture with delays were used with two hidden layers and 20 hidden nodes per layer. A hyperboilc tangent sigmoidal activation function was used and results were noted. The aim of the NN was to act as a controller that would regulate the RF gun. For this, the mean absolute error (MAE) and maximum errors were calculated. The NN showed a very tolerable, low MAE of *0.018* and the maximum error of *1.049*. Edelen et al. (2016) found the value setting time to be five times faster than the traditional method [26].

Overall, Edelen et al.(2016) were able to find significant success in implementing the feedforward NN as a experiment control with no manual human involvement. They were able to improve control time over the system by demonstrating shorter parameter setting time. The work could be validated further by using additional training data. As there was no mention of the size and shape of training and test data, it was hard to determine transferability of the results. The network could be updated to use reinforcement learning as a means to tune the intuition of the NN. By using the NN as a control, this study demonstrated the flexibility of NN applications outside of it's typical use cases.

### Jet Analysis

Pattern recognition is also crucial for jet analysis where a narrow beam of sub-atomic particles known as hadrons are produced as a result of high energy collisions [39]. But the identification and estimation of energy produced by such jets pose a challenge due to large background noise of low energy hadrons that get produced alongside rare jets. Dong and Gyulassy (1993) state that while conventional analysis techniques worked for proton-proton (pp) collisions, they failed for proton-nucleus (pA) collisions as a result of higher nuclear background. To address this, the authors studied the application of feed-forward neural networks for jet analysis. They demonstrated in their paper that a high-pass linear neural filter could be trained to allow for a bias-free estimator of jet energy. The high-pass linear neural filter additionally remained "nearly" bias free even in case of pA collisions with high nuclear background. They extended their study and showed that they could recover the underlying primitive jet distribution with a high degree of accuracy. Finally, their methodology also provided for a quantitative estimate of average energy loss - a significant metric, sought after by physicists [24].

Dong and Gyulassy (1993) described their architecture to have a simple high pass filter in the first layer of their neural network based on a threshold. Their second layer performed a sort on the remaining values and the final third layer estimated the jet energy. For their analysis however, they assumed physics threshold values to be fixed while study dictates that they should be variable to optimise results [43]. Monte Carlo event generators were used to produce a training sample. To achieve the physics goal of recovering the primitive jet distribution from the distorted results of the neural network, the authors deconvoluted the filtered out jet values in the second layer. They however found that deconvolution led to propagation of the error, which increased with increase in jet energy. They proposed settling for the error since it was very small. The final constrained deconvolution errors were *1%* to *7%* and the response remained within *10%* of the desired value even under high nuclear background. This was previously impossible to note with the traditional analysis techniques [24].

The authors were able to quantitatively demonstrate the ability of neural networks to perform jet analysis even with large nuclear background which traditional techniques failed to do so far. They were able to attain new physics information by deconvoluting the network to gather primordial jet distribution. This approach could pave way for origins of such particles. The study does fall short in terms of realistic data for training, which Monte Carlo simulations do not provide. The study used fixed threshold values that would ideally be variable with the training and finally, the study acknowledged the loss of information leading to deconvolution errors as a result of the measurement process.

Top-quarks are the heaviest of all observed elementary particles that were discovered and are still being researched at CERN [37]. Scientists are looking for ways by which a trigger mechanism or a control can tag jets that originated from top-quarks. Pearkes et al. (2017) presented a methodology for discriminating top-quark originated jets (signal) from all other flavours (background). They believed that CNNs might not be the ideal network for tagging top-quarks. This was because the images that capture such energy deposits had no identifiable features such as corners or edges that might aid in the learning process. It was determined that lack of identifiable features combined with sparse images would not result in the best performance from CNNs [35].

The authors cited domain knowledge as the main driving force behind all pre-processing. Vectors representing the constituents of the jets were generated. As part of data preprocessing, these vectors were scaled, rotated, flipped and finally ordered. Monte Carlo simulations were used by the authors to generate 3.75 million top-quark jets and and equal number of background noise. The dataset was split such that 80% was used for training the network. 10% of the remaining data was used for validation and another 10% for testing. For the (DNN), an input layer comprising of individual jet constituents was initialised. This was followed by 4 hidden layers and an output layer that presented a binary prediction. RELU was used as the activation function for the hidden layers and the sigmoid function was used for the output layer. An adaptive learning rate optimiser - Adam was used for training to handle sparse gradients and noise [35].

Receiver operating characteristic (ROC) curves were used to assess the ability of the DNN to reject background. ROC curve showed background rejection of *45* at *50%* efficiency operating point. Background rejection was *65* for the top-quark at the same efficiency. Area under the curve (AUC) reported *0.934* and *0.946* for background and top-quark respectively [35].

This study proved an interesting approach as the typical methodology of using images for particle tagging was rejected for a feed-forward neural network. The reported scores indicated that the network was able to distinguish between these two classes to a very high degree. The distinguishing aspect of this study was that information loss was at a minimum, since data was not converted to images. The authors did not compare obtained results against a benchmark. This prevents the ability to fairly judge if the DNN played a relevant role in the experiment. As the authors expressed a wish to maintain sequence processing, further work on this topic could be continued with Recurrent Neural Networks and Long Short Term Memory (LSTM) units [40].

Standard Model (SM) of physics describes known elements and their interactions with environment and with each other. However, there are particles that are beyond the Standard Model (BSM) and hence remain largely unexplained. Long lived particles (LLP) are one such type of particles from the BSM model that are being examined at the Large Hadron Collider (LHC) [28]. Calorimeters are part of experiments that try to determine the energy displaced by such particles from jets. In their study, Bhattacherjee et al. (2019) converted data captured by calorimeters to images. In these images, energy deposits tracked by the calorimeters were represented as 2D image intensities. The authors then trained a CNN to distinguish between LLPs and non-LLPs using 28x28 images of energy deposits. Non-displaced energy events were considered background noise and the displaced energy events were considered as LLP events. As such, four classes for each level of displacement were established. The network was adjusted to have two convolutional layers and non-linearity was introduced via RELU. L2 regularisation was applied to weights and the outputs were batch normalised. Each output layer was followed by a max-pooled layer that reduced the dimension of the image by half. Next, the output was flattened and passed through the softmax function. Adam was used as an optimiser with a learning rate of *0.001*. The network was trained on 60,000 images and validated and tested on 20,000 images respectively. A ROC curve was used to assess signal efficiency against background rejection.

For a signal efficiency of *60%*, they noted background rejection of *81.70%, 91.06%, 96.39%* and *99.6%* for four levels of displacement. The performance of the network was extremely high for the higher displacement cases. The authors Bhattacherjee et al. (2019) found these results to be significantly better than existing physics results averaging around *68%* across the four levels. They concluded that the CNN was able to learn energy displacement features from images and was able to discriminate between the background noise and LLPs [15] .

The study conducted by the authors has been the first of it's kind whereby energy was represented via image properties. The network achieved spectacular results on the generated images. It would be imperative to know how the model performs on real world data. The authors additionally do not describe any image pre-processing so it must be assumed that there was none, or based on physics reasoning. It would also be interesting to examine how the model performance changed based on the resolution of images. Moreover, it would be important to note that as spatial energy data was compressed into 2D images, there would have been been loss of information, which was not addressed in the study.

Particle physics experiments have several use cases of Field Programmable Gate Array (FGPA) based triggers and data acquisition systems because of the need for very low latency requirements by the detectors. Duarte et al. (2018) presented a case for neural networks in FGPA hardware. Such hardware is characterised by low power and low latency. The authors addressed the inference of DNNs for classification of jet substructures with FGPA hardware as either a quark, a gluon, a W or a Z boson or a top-quark jet. With current trigger strategies, these substructures are almost irretrievable amidst the background noise. The authors trained a fully connected DNN with three hidden layers. ReLU was used as the activation function for the hidden layers and softmax for the output layer. Adam was used to minimise the categorical cross-entropy loss function.

The neural network was able to classify top-quarks the best with *82%* accuracy. It successfully classified quarks, gluons, w and Z bosons with *73%, 76%, 74%* and *71%* accuracy respectively. Duarte et al. (2018) further adapted this neural network to work on FGPA hardware. To meet the constraints of the hardware, a compressed three layer neural network was recreated and made efficient. First the model was compressed via iterative parameter pruning and retraining with L1 regularisation. This led to *70%* of the parameters being pruned. Next, the model was quantized by reducing precision and pre-calculating non-trivial functions. The final compressed three layer model was run using FGPA hardware. It achieved a latency of *70 to 150 nanoseconds* and a clock frequency of *200 MHz*. These results lay well within the industry standard latency range of *1 microsecond* [25].

The study by Duarte et al. (2018) presented a very promising and interesting use case of neural network for classification of jet substructures. The authors were able to achieve accuracy upto *82%*. They furthered their application by adopting the DNN to FGPA hardware. They were able to run a compressed version of the same network with latency results well below the required standards. The paper does not discuss the use of more sophisticated neural networks such as recurrent neural networks and Long Short Term Memory (LSTM) units and how these could be adapted for FGPA. As the jet processes are often time derived, LSTM would be beneficial for analysis and importing it to the FGPA hardware might be a challenge.

**Energy Reconstruction**

Bai et al. (2016) propagated the use of a Bayesian neural network (BNN) to detect primary energies from interaction between cosmic rays and atmosphere. Extended Air Shower (EAS) arrays allow for observation of primary energy of a shower by evaluating the lateral density of shower particles and noting it's distance from the core. The authors Bai et al. (2016) used BNNs to estimate energy in EAS array experiments. As it was a non-linear function that they argued that BNNs could provide better control over model complexity. For their experiment, a toy detector array was designed and 3000 cosmic showers were simulated using SYBILL - a physics based event generator. The authors presented the energy reconstructions using a BNN with a single hidden layer and a linear fitting model (LFM) that served as a physics baseline. BNN was able to reconstruct energy with *28.2%* higher accuracy than the LFM at a lower levels of cosmic rays. The BNN was able to improve energy reconstruction at the rate of *43.0%* over the linear model at extremely high levels of cosmic rays [11].

The study was able to successfully implement a neural network based alternative with significant improvements in results over the baseline analysis. However, the baseline technique was modelled on a linear relationship. It would be interesting to compare a baseline non-linear function with the non-linear BNN for a fairer, equivalent comparison.

**Neutrinos**

Szadkowski et al. (2014) proposed extending neural networks (NNs) to study air showers that resulted from neutrinos interacting with the Earth's atmosphere (down-going) or crust (Earth-skimming) on behalf of the Pierre Auger observatory. Amidst background noise of cosmic rays, detecting the very infrequently occurring neutrino showers has been the main challenge for the observatory. All types of neutrino showers were categorised into two event groups based on their electromagnetic properties - young showers and old showers. Szadkowski et al. (2014) stated the inability of the

Pierre Auger trigger to detect young neutrino showers thus far and hypothesised it being due to the fact that the current triggers were not sufficiently sensitive [43].

The authors Szadkowski et al. (2014) first presented a trigger based on Discrete Cosine Transform (DCT) that could perform pattern recognition to identify traces corresponding to the old showers associated with both down-going and Earth-skimming [42] events. DCT is a technique that is used to separate an image into sub-parts or bands where each band has differing importance with respect to the image quality [6]. Thus, it can transform image from spatial or time domain to frequency domain. The DCT trigger was able to detect all old showers based on sample test data [43].

The authors next proposed a NN for for the harder to detect young showers. They proposed use of the Levenberg-Marquardt algorithm for backpropagation. The Levenberg-Marquardt algorithm or the Damped Least Squares (DLS) is encouraged when sums of squares of non linear functions need to be minimised [33]. Computationally, the Levenberg-Marquardt was designed to allow the computation to increase speed without having to compute a Hessian matrix [33]. The NN with the Levenberg-Marquardt backpropogation algorithm was set up to identify both young and old showers using simulated Monte Carlo events. A three layer NN with hyperbolic sigmoid transfer function (*tansig*) was trained on 245,760 different patterns grouped as 160 events. The authors presented extremely promising results. Noise was perfectly rejected and the NN was able to to identify 161 patters out of the 160, with one false-positive. Thus on simulated data, the authors showed the ability of the NN to detect young showers with a very low error rates that the Auger observatory was unable to detect [43].

While Szadkowski et al. (2014) presented the first known implementation of NNs on neutrino data, the paper faced a few shortcomings. First, the authors did not rationalise their choice of hyperparameters and the reason for using a three layer network. Further, the authors only talked about error rates but did not discuss a few other metrics that could allow the user to assess the performance of the NN.

Liquid Argon time-projection chambers (LArTPCs) are a kind of particle detectors that produce high resolution images of particle interactions. Acciarri et al. (2016) examined deep convolutional neural networks to reconstruct neutrino scattering interactions in LArTPCs. For this, neutrino scatterings within captured images had to be first identified and then classified, which justified the use of CNNs. They explored the use of CNNs for detector images that were very information sparse and contained lines from particle paths that were far apart from each other. Thus, resulting images were often empty. Acciarri et al. (2016) investigated the ability of CNNs to classify images of single particles located in various parts of the detector. They found that CNNs could detect these particles to varying degrees and represent them using bounding boxes [3].

22,000 events per type of particle were used for training the CNN in batches. Both high (576 by 576 pixels) and low resolution images (288 by 288 pixels) were provided as separate demonstrations to mimic realistic scenarios. The authors suggested choosing batch size based on network size and GPU limitations. AlexNet was trained for particle classification and loss was minimised using stochastic gradient descent (SGD). Next, for particle detection, Faster-RCNN was trained to localise objects within images. Faster-RCNN was provided with training images and equivalent truth labels and ground truth bounding box. Faster-RCNN returned N classification predictions and a minimum bounding box area. The Faster-RCNN segment of the network was combined with AlexNet after it's fifth convolutional layer. This allowed for combined localisation and particle classification on images [3].

The results very quite promising as the authors noted R-CNN combined AlexNet's ability to distinguish track-like particles from shower particles very well. For high resolution images, track-like particles had *87.2%* accuracy and *81.3%* accuracy for shower-like particles. For low resolution images, the score was understandable lower with *85.8%* accuracy for track-like images and *77.3%* for shower-like images [3].

Based on the accuracy scores, it can be concluded that there was reasonable localisation of both shower and track particles for high and low resolution images. It would be imperative to note that shower-like particles had the highest likelihood of being wrongly classified for both high and low resolution images. Based on the setup, the customised architecture might not generalise well outside of the simulated environment that was set up by the authors. But the key point was that the authors were able to demonstrate the effectiveness of CNNs even with sparse images.

Neutrino event classification experiments at the LArTPC chambers also involve tagging and identification of on-beam event images for a neutrino interaction. Authors Acciarri et al. (2016) developed a methodology that identified neutrino interactions on single plane images and cropped them around the interaction region. They then applied the network they described in their previous work to classify particles in the cropped images [2, 3].

For training, Acciarri et al. (2016) generated Monte Carlo images where an equal number of neutrino events were overlaid with cosmic background images from off-beam events. A total of 101,191 images were thus generated for training (and 32,220 images were used for validation). Two classes were defined for the classification task - Monte Carlo neutrino events and purely cosmic events. InceptionResNet, composed of three different modules was used for the task. Since detector images tend to be larger and of higher resolution, the modules per block were reduced to allow for it to meet memory constraints. FasterRCNN and AlexNet were once again used as in their previous study to train for neutrino particle detection [3]. As part of data preparation, random cropping of images were performed for each time the image was given to the network. The authors reported an *80%* accuracy score during training but faced certain amount of over-training indicated from lower validation scores (*78%*). The authors deemed this acceptable due to the large number of parameters. Next for detection training, Faster-RCNN and AlexNet were once again trained but with modifications to allow the output to be two classes - neutrino or background. The authors re-initialised the last fully connected layer with Gaussian weights instead of recycling weights used at the classification state. They justified that this allowed for detection-specific layers to learn both bounding-box regression and classification. Performance was stated as very positive with *80.1%* selection efficiency for neutrino events. They believed that this efficiency would improve if all three planes were used instead of just one plane (as in this study) [3].

Authors Acciarri et al. (2017) continued on the study from [2] by extending their work from single-plane images to three-planes and combining it with optical detector data [21]. While maintaining the same strategy of using simulated neutrino images and cosmic images (as background), the input images were left at 768 by 768 pixels and 12 channels as the third dimension. This resulted in a large amount of data. A new truncated network based on ResNet was designed where the network repeatedly used residual convolutional modules for faster training. The authors discussed the compromise of having fewer layers learning fewer filters but preserving resolution, allowing for exposure to detailed features. The three planes were passed individually through thee convolutional layers and pooling layers to reduce the size of the feature map. In addition to the images, the authors provided three supporting images for each plane as additional information. Distribution of neutrino classification scores showed a very good separation between the two types of events. The selection efficiency improved to *85%* (from *80.1%* with one plane images) [21].

Authors Authors Acciarri et al. (2016, 2017) [2] [3] [21] showed promising demonstrations of particle classification, particle and neutrino detection and neutrino event identification, all using forms of CNNs. However, it is unknown if these metrics can be maintained for real world data. It is also unclear if similar high results can be obtained by generalising these models to data from other particle detectors. Biases in network learning might be possible due to the use of simulated images. The authors however made good use of several combined architectures and provided a good starting point for neutrino research for events in LArTPC. They also developed an architecture that could be potentially included as part of the detector pipeline.

Adams et al. (2018) continued on the work of Acciari et al. (2016, 2017) by developing a convolutional neural network that could predict objects in image data at the pixel level [22]. They also built their model using data captured by LArTPC and demonstrated that electromagnetic particles (EM) could be discriminated from others at the pixel level using CNNs [22].

Adams et al. (2018) trained U-ResNet, a deep semantic segmentation network via supervised learning. First they used transfer learning techniques by training the first half of the network on the dataset from a previous work that contained single particle images [3]. Then they developed a new loss factor called pixel-wise loss (PL) weighing factor. This factor was multiplied by a single pixel's loss contribution to the total loss of an image. Thus, challenging sections of the image obtained higher weighted pixel loss, enabling the network to focus it's training on those regions. Following the loss weighting procedure, RMSProp was used to optimise U-ResNet and the process was monitored using the Incorrectly Classified Pixel Fraction (ICPF) metric. The ICPF mean scored indicates the average value of incorrectly classified pixel per image over all images on all events in a sample [46]. The network was trained on 100,000 images and then tested on 20,000 images. U-ResNet achieved an average ICPF of *6.0* for electron neutrinos and *3.9* for muon neutrinos. They noted that U-ResNet could classify pixels from low energy and simple topologies fairly well with mean ICPF scores of *2.3* and *3.9* respectively. The low mean ICPF scores once again demonstrated the ability of CNNs to work with neutrino particles at an even deeper, pixel level. The authors additionally obtained real detector data called Michel electron data and compared the results of their network with those obtained by physicists. They found that the physicists had a lower mean ICPF score of *1.8* for the electron samples while the network scored a mean ICPF of *2.6* [22].

The physicists had better results when compared to the network and the authors believed this to be because the network focused on physics features in the image. They believed that addition of specialised, handcrafted features driven by physics knowledge might resolve the differences. Despite this, the study should be considered informative since the network did not have significantly poor performance, whilst exploring a new methodology for training. The authors were convinced that once the gap between the two metrics was closed, their NNs would be a suitable candidate for the detector's data reconstruction pipeline.

IceCube is a neutrino observatory at the South Pole that solely searches for high energy neutrino events [20]. It observes two classes of such events - neutrino interactions within the detector and high energy cosmic interactions in the upper atmosphere [20]. The detectors are physically arranged in an irregular shape and faces sparse signals [19].

Choma et al.(2018) in their study on data from IceCube proposed that the irregular geometry of the detectors can be modelled as a graph with vertices as sensors and edges as learned functions if the sensors spatial coordinates. They stated a large asymmetry between positive and negative events as the main challenge. The authors proposed the use of Graphical Neural Networks (GNNs) for this work. GNNs were further deemed suitable since the IceCube detector array is hexagonal and irregular eliminating the assumption of stationarity. GNNs do not require such an assumption. The authors generated two Monte Carlo datasets to represent signal and background data. They considered muon neutrinos as positive signals and the rest as negative background. As the background was much larger in terms of magnitude than the signal, a high rejection power was required. The GNN was initialised as a fixed, weighted, directed graph. Output features of the last convolutional layer were pooled and passed through the sigmoid function. 25,250 events were generated as signal and 109,491 events were generated as background with 50% being used for training from both datasets and 25% and 25% used for validation and testing. The performance of the classification was noted against physics results and CNN scores were used as baseline [19].

As per Choma et al. (2018), physics-derived metrics reported *0.987* signal to noise ratio for events per year and CNN reported *0.937* signal to noise ratio. The GNNs showed clearly superior results by reporting *2.980* signal to noise ratio for events per year. The GNN outperformed physics metrics by identifying three times more signal (positive) events [19].

NOvA experiment aims to make precise measurements of neutrino oscillation parameters [34]. This requires reconstruction of neutrino energy and flavour. Aurisano et al (2016) developed a technique called Convolutional Visual Network (CVN) based on CNNs to achieve such goals. It was inspired by GoogLeNet architecture that uses network-in-network (NIN) methodology to reduce dimensionality and modify the learning capacity of convolutional layers. The authors differed from GoogLeNet in that they had two distinct views of the same image, rather than representing the single image in multiple colour channels. They also cut short the GoogLeNet network after three inception modules on account of having simpler images. Output from the final module was down-sampled using an average pooling layer and classifier outputs were calculated using softmax or exponential function during forward pass. The network was trained using mini-batches on 3.7 million simulated neutrino events and tested on 1 million samples [10].

The authors Aurisano et al. (2019) noted that they were able to obtain optimal convergence by dropping the step size of SGD at fixed intervals. Additional regularisation techniques were maintained such as adding penalty terms to back-propagation calculations. To measure the CVN's performance, it was first compared against existing metrics. Measurement-optimised efficiency scores were obtained from existing physics metrics and compared against that of the CVN. CVN scored an efficiency of *58%* versus the existing *57%* efficiency for muon neutrino interactions. The authors felt that while the improvement was modest, it was still in the positive direction. CVN however scored *40%* efficiency over the pre-existing metric of *35%* for electron neutrinos. Additionally, the authors computed a Figure of Merit (FOM) to assess the performance of signal identification over background noise for oscillation parameters. Overall, the CVN obtained a range of efficiency scores from *17.4%* at the lowest to *66.4%* at the highest for various parameters. The authors found the results quite promising since they performed minimal event reconstruction and found positive performance with a single algorithm [10].

Moreover, the CVN developed was used on atypical images, specifically the readout of a calorimeter. This study therefore opens up the possibility of using a different medium of data which might be applicable to other detectors as well.

## 4 DISCUSSION

The studies reviewed thus far have all showed the nature of work in particle physics and the ways in which neural networks could be used to fill the gaps. Most analysis aimed at discovering new particles involves distinguishing between signal and background noise. With limited data and expensive infrastructure required to study such particles, improvements to these physical tools are quite constrained. This opens up other avenues for improvement - neural networks [10]. The studies explored thus far all agreed with this sentiment and applied neural networks to data from various particle detectors to answer different questions. Table 4 summarises the results by showing the attempted discovery and neural network used for the study.

| **Reference** | **Discovery** | **Type of NN** |
|---|---|---|
| Baldi et al. (2014) | Higgs Bosons & supersymmetric particle | Feedforward NN (5 layer) |
| Szadkowski et al. (2014) | Young and old Neutrino Showers | Feedforward NN (3 layer) |
| Edelen et al. (2016) | Resonant Frequency Gun Control | Feedforward NN (2 layer) |
| Dong & Gyulassy (1993) | Jet Energy & Primordial distribution | Deconvolutional CNN |
| Pearkes et al. (2017) | Top-quarks in Jets | Feed Forward NN (4 layers) |
| Bhattacherjee et al. (2019) | Long Lived Particle Jet Energy | CNN |
| Duarte et al. (2018) | Jet Subsctructures | Feedforward NN (3 layer) |
| Bai et al. (2016) | Cosmic ray energy reconstruction | Bayesian NN |
| Acciarri et al. (2016) | Neutrino particle | RCNN + AlexNet |
| Acciarri et al. (2016) | single plane Neutrino particle | FasterRCNN & AlexNet |
| Acciarri et al. (2017) | On-beam Neutrino particle | ResNet |
| Adams et al. (2018) | Pixel level neutrino flavours | U-ResNet |
| Choma et al. (2018) | Neutrino flavours | Graphical NN |
| Racah et al. (2016) | Anti-neutrino particles | CNN & Autoencoder |
| Aurisano et al. (2016) | Neutrino flavours | Convolutional Visual Network |

**Table 4: Summary of Results**

The neural networks trained were all more or less straightforward and involved feedforward networks and convolutional neural networks with variations in hyperparameters and layers [12, 26, 43]. Often, studies converted data to images and applied CNNs. The images passed on to the CNNs all shared similar characteristics of being very large, very sparse and of high quality. However, studies often times left out justification for choosing hyperparameter values [12, 43]. Knowing their reasoning would be useful to assess soundness of the values and thus fully understand the results. Aside from CNNs and feedforward networks, a study on cosmic rays used Bayesian Neural Network to estimate energies from cosmic air showers [11]. Some studies combined a few architectures such as FasterRCNN with AlexNet, to create an informal pipeline that performed a set of relevant tasks [2, 3]. One study in particular chose to use Graphical Neural Networks (GNNs) instead of CNNs based on the shape of the initial dataset [19]. With this in mind, the study reported higher scores over CNNs and existing physics metrics that were run on the same dataset [19]. A study on anti-neutrino particle identification additionally used unsupervised learning with the help of a convolutional autoencoder [36]. This revealed that unsupervised learning was a suitable candidate for particle classification tasks. If unsupervised learning could be adopted more widely, the overall experimentation process would be simplified with less need to generate complex, handcrafted features. But additional research would be required to assert this.

Most of the problems discussed involved identification of particles from background noise and then classification based on subfields of particle physics. Under sub-atomic particle study, studies used neural networks to classify Higgs Bosons and supersymmetric particles from the background noise [12]. A few other studies focused on jet analysis of sub-atomic particles from collisions. Such analysis included measuring energy and tracing the particles primordial distribution [24]. A unique methodology was tested where energy displacement measurements by a calorimeter were represented as intensities on 2D images and passed onto CNNs [15]. Also under jet analysis, a neural network was trained to additionally classify jet substructures on restrained hardware [25] Other neutrino focused studies classified neutrino showers into young and old showers [43]. A study at the Fermilab Accelerator successfully used neural networks to control a physical equipment at the accelerator which presented an alternate and unique use case for neural networks [26]. Another study accounted for the nature of hardware prevalent in the detectors and presented a network that was specifically adapted to run on them [25]. Notably the first of its kind, the study set a precedent for future work that attempts to improve networks being run on such hardware. The remainder of the studies focused on working with neutrino data. Several studies used experiment data from Liquid Argon Time Projection Chambers (LArTPC) to identify particles in images (localisation) and then classify them to various degrees [2, 3]. These studies additionally compared results between high and low resolution images to provide a comprehensive analysis. Other neutrino studies focused on separating neutrino interactions from cosmic interactions [19]. Neutrino oscillation parameters and anti-neutrino particle detection were also examined with data from various detectors [10, 36]. Additionally, for the work on neutrino oscillation parameters, the authors trained their network to read from instrument readouts rather than particle or space images [10]. This also has added potential for expanding the scope of what CNNs can be trained on.

In several cases such as jet analysis, where data was time sensitive, network architecture could be extended to include Long Short Term Memory (LSTM) units, for more intuitive learning [24, 25].

Different applications call for different representations of the metrics. For example, some general classification related metrics include reporting on accuracy, precision, recall, F1-scores, ROC and AUC [32]. Due to the varied nature of the datasets and questions being answered, a varied number of metrics got reported. This made it challenging to compare the successes of studies with respect to one another. There were very few studies that stated comparisons of their networks against the standard physics results. Baldi et al. (2014) pointed out an interesting metric - discovery significance in their study [12]. This could be adopted by future studies assessing neural networks for their field as it presents an interesting perspective of the potential for discovery. Often there was no reporting on loss and the extent to which it was minimised by the study.

All of the studies did not leave much scope for being generalised across other detectors and experiments based on the described method. Moreover, the training data itself was more often simulated or toy data thus leaving open questions regarding its extensibility to real world data [2, 3, 21, 22]. Not many of these studies discussed production value potential of being incorporated as part of the larger detector system. Finally, there was no talk about implications on hardware being used or the performance or training hours. It is possible that these were considered insignificant as the population that conducted these studies themselves used high performance computing systems that is often available to such particle detectors and based on the details outlined in the report, did not engaging in online training or training with massive samples. This is something that must be kept in mind by readers attempting to recreate some of these experiments.

# 5 CONCLUSION

The nature of particle physics research is such that most of the work is often exploratory and researchers usually do not know what they should be looking out for. Particle detectors are often the source for studying and collecting data. These instruments gather exabytes of data that need to be carefully probed from all angles to make new discoveries. While the state-of-art in instrument and hardware has significantly improved, this has allowed for detailed data capture. Traditional physics algorithms and controls are unable to keep up with the burst of high quality, irregular data, making new discoveries harder and enforcing more manual labour on the researchers themselves. With this gap between inferior information extraction techniques and superior information gathering systems, researchers can look to artificial intelligence to fill the gap. While neural networks have not been adopted as widely, there have been some attempts to adopt them into research pipelines. This could be credited to new advances in DNNs themselves such as having the ability to train with multiple hidden layers, improved speedup of stochastic gradient descent algorithms using graphics processors, new learning algorithms and introduction of methods such as autoencoding.

The study set out to gather and summarise all the work that has been done so far in the realm of particle physics using neural networks. Doing so would serve as an overview of the state-of-the-art in neural networks under this branch of physics and act as a reference for those in the field wishing to adopt neural networks themselves. Summaries of the methods undertaken, the problems addressed and the results obtained could serve as a starting point for others in particle physics wishing to adopt neural networks to their own work but not necessarily having technical expertise in the area of artificial intelligence.

A search strategy was adopted to identify thirty three studies that were then filtered down to seventeen studies based on an Inclusion/Exclusion criteria. Referring back to the research questions that were identified in Section 2.2, it was seen that much of the work in particle physics required classification of particles, separating signals from copious amounts of background noise or pattern recognition and inferring secondary properties of particles (*RQ-I.I*). Convolutional neural networks (CNNs) were the most commonly adopted networks for these tasks, followed by feedforward perceptrons (*RQ-I*). Often the resulting data were treated as images that CNNs are known to work well with. However, the sparse nature of these images were different from the typical images CNNs get trained on. Studies showed an overall preference for using simpler architectures with very few studies attempting to build more customised pipelines. This could be attributed to the fact that the researchers carrying out this study had a physics background. Overall, despite very few tweaks to the original structure, CNNs were able to do the job of classification well. Feed-forward perceptrons were noted to be the other popular choice, likely due to it's simpler set up. There was a varied use of metrics for diagnosis and performance measurements for similar tasks across different studies. Out of the studies that compared their results with existing physics metrics or a certain baseline, it was seen that all neural networks performed better (*RQ-II*). The CVN developed by Aurisano et al.

(2016) scored slightly below the available physics metric but the decrease was considered insignificant [10].

Based on the successes of the described papers, this study can conclusively state that CNNs would be an ideal starting point for any similar exploratory work with particles and images as data. However, some of the studies discussed converted data to images in order to use CNNs. This indicates a potential for inefficiency and loss of information. Additional networks and models need to be explored to determine if they can be used for the data without any intermediate conversion of data to images. The role of recurrent or modular neural network architectures could be explored for example. The nature of particle physics is such that often variables and their relationships with each other are less understood. Thus, caution should be used when employing such variables with neural networks that have not had any prior adjustments. Further, caution should be used when interpreting results and checks should be in place to ensure that overfitting is minimised. Most of the training was performed with simulated data, rather than existing data collected by detectors. It would be advised to use real world data as much as possible. Experiments should also try to assess results against current physics standards to quantitatively determine the benefits of using neural networks. Finally, this study recommends that future work also assess feasibility of adopting the networks into a full development pipeline such that it can be truly incorporated into the system and made an official part of ongoing research.

The study shows that there is certainly more in-depth work required to understand how complex networks could be applied to the field. Additional studies would be required to investigate data preparation and pre-processing by itself. Work could be focused on developing architectures that specialise on training with extremely sparse images. With these gaps in place however, neural networks at their current state have a very promising role in the future of particle physics and by extension neutrino research.

## REFERENCES

[1] Halina Abramowicz, Allen Caldwell, and Ralph Sinkus. 1995. Neural network based electron identification in the ZEUS calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 365, 2-3 (1995), 508–517.

[2] R Acciarri, MA Acero, M Adamowski, C Adams, P Adamson, S Adhikari, Z Ahmad, CH Albright, T Alion, E Amador, et al. 2016. Long-baseline neutrino facility (LBNF) and deep underground neutrino experiment (DUNE) conceptual design report, volume 4 the DUNE detectors at LBNF. *arXiv preprint arXiv:1601.02984* (2016).

[3] R Acciarri, C Adams, R An, J Asaadi, M Auger, L Bagby, B Baller, G Barr, M Bass, F Bay, et al. 2017. Convolutional neural networks applied to neutrino events in a liquid argon time projection chamber. *Journal of instrumentation* 12, 03 (2017), P03011.

[4] P Adamson, C Andreopoulos, KE Arms, R Armstrong, DJ Auty, S Avvakumov, DS Ayres, B Baller, B Barish, PD Barnes Jr, et al. 2008. Study of muon neutrino disappearance using the Fermilab Main Injector neutrino beam. *Physical Review D* 77, 7 (2008), 072002.

[5] AA Aguilar-Arevalo, M Backfish, A Bashyal, B Batell, BC Brown, R Carr, A Chatterjee, RL Cooper, P Deniverville, R Dharmapalan, et al. 2017. Dark matter search in a proton beam dump with MiniBooNE. *Physical review letters* 118, 22 (2017), 221803.

[6] Nasir Ahmed, T_ Natarajan, and Kamisetty R Rao. 1974. Discrete cosine transform. *IEEE transactions on Computers* 100, 1 (1974), 90–93.

[7] Sebastiano Aiello, SE Akrame, F Ameli, EG Anassontzis, Michel Andre, G Androulakis, Marco Anghinolfi, G Anton, Miguel Ardid, Julien Aublin, et al. 2019. Sensitivity of the KM3NeT/ARCA neutrino telescope to point-like neutrino sources. *Astroparticle Physics* 111 (2019), 100–110.

[8] A Albert, M André, M Anghinolfi, G Anton, M Ardid, J-J Aubert, J Aublin, B Baret, S Basa, B Belhorma, et al. 2020. ANTARES and IceCube Combined Search for Neutrino Point-like and Extended Sources in the Southern Sky. *arXiv preprint arXiv:2001.04412* (2020).

[9] FP An, JZ Bai, AB Balantekin, HR Band, D Beavis, W Beriguete, M Bishai, S Blyth, RL Brown, I Butorov, et al. 2016. The detector system of the Daya Bay reactor neutrino experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 811 (2016), 133–161.

[10] A Aurisano, A Radovic, D Rocco, A Himmel, MD Messier, E Niner, G Pawloski, F Psihas, Alexandre Sousa, and P Vahle. 2016. A convolutional neural network neutrino event classifier. *Journal of Instrumentation* 11, 09 (2016), P09001.

[11] Ying Bai, Ye Xu, Jia Pan, JieQin Lan, and WeiWei Gao. 2016. Application of Bayesian neural networks to energy reconstruction in EAS experiments for ground-based TeV astrophysics. *Journal of Instrumentation* 11, 07 (2016), P07006.

[12] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. 2014. Searching for exotic particles in high-energy physics with deep learning. *Nature communications* 5 (2014), 4308.

[13] Jacob D Bekenstein. 1973. Black holes and entropy. *Physical Review D* 7, 8 (1973), 2333.

[14] R Bellotti, M Castellano, C De Marzo, N Giglietto, G Pasquariello, and P Spinelli. 1993. A comparison between a neural network and the likelihood method to evaluate the performance of a transition radiation detector. *Computer physics communications* 78, 1-2 (1993), 17–22.

[15] Biplob Bhattacherjee, Swagata Mukherjee, and Rhitaja Sengupta. 2019. Study of energy deposition patterns in hadron calorimeter for prompt and displaced jets using convolutional neural network. *Journal of High Energy Physics* 2019, 11 (2019), 156.

[16] SM Bilenky. 2013. Neutrino. History of a unique particle. *The European Physical Journal H* 38, 3 (2013), 345–404.

[17] Léon Bottou. 2010. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*. Springer, 177–186.

[18] Serguei Chatrchyan, V Khachatryan, AM Sirunyan, W Adam, B Arnold, H Bergauer, T Bergauer, M Dragicevic, M Eichberger, J Ero, et al. 2010. Performance of the CMS Level-1 trigger during commissioning with cosmic ray muons and LHC beams. *Journal of Instrumentation* 5 (2010).

[19] Nicholas Choma, Federico Monti, Lisa Gerhardt, Tomasz Palczewski, Zahra Ronaghi, Prabhat Prabhat, Wahid Bhimji, Michael Bronstein, Spencer Klein, and Joan Bruna. 2018. Graph neural networks for icecube signal classification. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 386–391.

[20] IceCube Collaboration et al. 2013. Evidence for high-energy extraterrestrial neutrinos at the IceCube detector. *Science* 342, 6161 (2013), 1242856.

[21] MicroBooNE Collaboration. 2016. Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber. *arXiv preprint arXiv:1611.05531* (2016).

[22] MicroBooNE Collaboration, C Adams, M Alrashed, R An, J Anthony, J Asaadi, A Ashkenazi, M Auger, S Balasubramanian, B Baller, et al. 2019. Deep neural network for pixel-level electromagnetic particle identification in the MicroBooNE liquid argon time projection chamber. *Physical Review D* 99, 9 (2019), 092001.

[23] Bruce Denby. 1999. Neural networks in high energy physics: a ten year perspective. *Computer Physics Communications* 119, 2-3 (1999), 219–231.

[24] Dawei W Dong and Miklos Gyulassy. 1993. Neural filters for jet analysis. *Physical Review E* 47, 4 (1993), 2913.

[25] Javier Duarte, Song Han, Philip Harris, Sergo Jindariani, Edward Kreinar, Benjamin Kreis, Jennifer Ngadiuba, Maurizio Pierini, Ryan Rivera, Nhan Tran, et al. 2018. Fast inference of deep neural networks in FPGAs for particle physics. *Journal of Instrumentation* 13, 07 (2018), P07027.

[26] AL Edelen, SG Biedron, BE Chase, D Edstrom, SV Milton, and P Stabile. 2016. Neural networks for modeling and control of particle accelerators. *IEEE Transactions on Nuclear Science* 63, 2 (2016), 878–897.

[27] FE Gray, C Ruybal, J Totushek, D-M Mei, K Thomas, and C Zhang. 2011. Cosmic ray muon flux at the Sanford Underground Laboratory at Homestake. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 638, 1 (2011), 63–66.

[28] Howard E Haber and Gordon L Kane. 1985. The search for supersymmetry: probing physics beyond the standard model. *Physics Reports* 117, 2-4 (1985), 75–263.

[29] Jonathan J Halliwell and Stephen William Hawking. 1985. Origin of structure in the universe. *Physical Review D* 31, 8 (1985), 1777.

[30] Siddhartha R. Jonnalagadda, Pawan Goyal, and Mark D. Huffman. 2015. Automating data extraction in systematic reviews: a systematic review. *Systematic Reviews* 4 (2015).

[31] Barbara Kitchenham and Stuart Charters. 2007. Guidelines for performing systematic literature reviews in software engineering. (2007).

[32] Oluwasanmi O Koyejo, Nagarajan Natarajan, Pradeep K Ravikumar, and Inderjit S Dhillon. 2014. Consistent binary classification with generalized performance metrics. In *Advances in Neural Information Processing Systems*. 2744–2752.

[33] Chen Lv, Yang Xing, Junzhi Zhang, Xiaoxiang Na, Yutong Li, Teng Liu, Dongpu Cao, and Fei-Yue Wang. 2017. Levenberg–Marquardt backpropagation training of multilayer neural networks for state estimation of a safety-critical cyber-physical

system. *IEEE Transactions on Industrial Informatics* 14, 8 (2017), 3436–3446.

[34] RB Patterson. 2012. The NOvA experiment: status and outlook. *arXiv preprint arXiv:1209.0716* (2012).

[35] Jannicke Pearkes, Wojciech Fedorko, Alison Lister, and Colin Gay. 2017. Jet constituents for deep neural network based top quark tagging. *arXiv preprint arXiv:1704.02124* (2017).

[36] Evan Racah, Seyoon Ko, Peter Sadowski, Wahid Bhimji, Craig Tull, Sang-Yun Oh, Pierre Baldi, et al. 2016. Revealing fundamental physics from the Daya Bay Neutrino Experiment using deep neural networks. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 892–897.

[37] Keith Rehermann and Brock Tweedie. 2011. Efficient identification of boosted semileptonic top quarks at the LHC. *Journal of High Energy Physics* 2011, 3 (2011), 59.

[38] Arthur Roberts. 1992. The birth of high-energy neutrino astronomy: A personal history of the DUMAND project. *Reviews of Modern Physics* 64, 1 (1992), 259.

[39] Craig D Roberts, MS Bhagwat, A Höll, and SV Wright. 2007. Aspects of hadron physics. *The European Physical Journal Special Topics* 140, 1 (2007), 53–116.

[40] Hasim Sak, Andrew W Senior, and Françoise Beaufays. 2014. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. (2014).

[41] Stefan Seuring and Martin Müller. 2008. From a literature review to a conceptual framework for sustainable supply chain management. *Journal of cleaner production* 16, 15 (2008), 1699–1710.

[42] Z Szadkowski. 2009. A spectral 1st level FPGA trigger for detection of very inclined showers based on a 16-point Discrete Cosine Transform for the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 606, 3 (2009), 330–343.

[43] Zbigniew Szadkowski and Krzysztof Pytel. 2015. Artificial neural network as a FPGA trigger for a detection of very inclined air showers. *IEEE Transactions on Nuclear Science* 62, 3 (2015), 1002–1009.

[44] Masahiro Tanaka, K Abe, C Bronner, Y Hayato, M Ikeda, S Imaizumi, H Ito, J Kameda, Y Kataoka, Y Kato, et al. 2020. Search for proton decay into three charged leptons in 0.37 megaton-years exposure of the Super-Kamiokande. *Physical Review D* 101, 5 (2020), 052011.

[45] GA Voss and BH Wiik. 1994. The electron-proton collider HERA. *Annual Review of Nuclear and Particle Science* 44, 1 (1994), 413–452.

[46] Liangpei Zhang, Ke Wu, Yanfei Zhong, and Pingxiang Li. 2008. A new sub-pixel mapping algorithm based on a BP neural network with an observation model. *Neurocomputing* 71, 10-12 (2008), 2046–2054.