NATIONAL TECHNICAL
UNIVERSITY OF ATHENS

SCHOOL OF APPLIED
MATHEMATICAL
AND PHYSICAL SCIENCES

SCHOOL OF MECHANICAL
ENGINEERING

NCSR «DEMOKRITOS»

INSTITUTE OF NANOSCIENCE
AND NANOTECHNOLOGY

INSTITUTE OF NUCLEAR AND
PARTICLE PHYSICS

«Physics & Technological Applications»

Event Reconstruction using Graph Neural Networks for the KM3NeT/ARCA8 detector

MASTER THESIS of
ELENI ANDROUTSOU

Supervisor: EVANGELIA DRAKOPOULOU

Athens, September, 2024

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΕΚΕΦΕ «ΔΗΜΟΚΡΙΤΟΣ»

ΙΝΣΤΙΤΟΥΤΟ ΝΑΝΟΕΠΙΣΤΗΜΗΣ ΚΑΙ ΝΑΝΟΤΕΧΝΟΛΟΓΙΑΣ

ΙΝΣΤΙΤΟΥΤΟ ΠΥΡΗΝΙΚΗΣ ΚΑΙ ΣΩΜΑΤΙΔΙΑΚΗΣ ΦΥΣΙΚΗΣ

Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
«Φυσική και Τεχνολογικές Εφαρμογές»

Ανακατασκευή γεγονότων με τον ανιχνευτή KM3NeT/ARCA8 με χρήση αλγορίθμων γράφων

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
της ΕΛΕΝΗΣ ΑΝΔΡΟΥΤΣΟΥ

Επιβλέπουσα: ΕΥΑΓΓΕΛΙΑ ΔΡΑΚΟΠΟΥΛΟΥ

Αθήνα, Σεπτέμβριος, 2024

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμότερες ευχαριστίες μου στην επιβλέπουσα της διπλωματικής μου εργασίας, Ερευνήτρια Δρ. Ευαγγελία Δρακοπούλου, η οποία μου έδωσε την πολύτιμη ευκαιρία να εκπονήσω την εργασία μου στο Ινστιτούτο Πυρηνικής και Σωματιδιακής Φυσικής του Ε.Κ.Ε.Φ.Ε. ¨Δημόκριτος¨. Η συνεχής υποστήριξη και καθοδήγησή της καθ΄ όλη τη διάρκεια αυτής της διαδρομής υπήρξαν καθοριστικές για την επιτυχή ολοκλήρωση της έρευνάς μου.

Επιπλέον, θα ήθελα να ευχαριστήσω τα υπόλοιπα μέλη της ερευνητικής ομάδας του πειράματος KM3NeT, για τη συνεισφορά τους και τις πολύτιμες συμβουλές τους. Χωρίς την καθοδήγηση και την τεχνική υποστήριξή τους, η εργασία αυτή δε θα μπορούσε να ολοκληρωθεί με επιτυχία.

Ένα ιδιαίτερο ευχαριστώ στο ίδρυμα Ωνάση για την εμπιστοσύνη και υποστήριξη που μου έδειξε, προσφέροντάς μου υποτροφία ώστε να περατωθούν επιτυχώς οι προπτυχιακές μου σπουδές.

Τέλος, ένα μεγάλο ευχαριστώ οφείλω στην οικογένειά μου και στους φίλους μου για τη συνεχή υποστήριξη, την ενθάρρυνση και την αμέριστη συμπαράστασή τους καθ΄ όλη τη διάρκεια των σπουδών μου.

Abstract

Neutrinos are subatomic particles of high interest for many fields of scientific research. Neutrinos, being electrically neutral, offer a unique tool to investigate the inner regions of astrophysical objects and to comprehend the Universe across a broader spectrum of energies. Their trajectory remains unaffected by magnetic fields, keeping directionality when reaching us from distant sources. There are various astrophysical sources known to emit neutrinos providing a direct means of observing these sources. Furthermore, evidence for a non-zero mass enabled studies on neutrino oscillations, contributing to our understanding of fundamental particle physics.

The KM3NeT water Čerenkov neutrino telescopes are optimized for the detection of high-energy neutrinos, in the GeV-PeV range. KM3NeT/ARCA neutrino detector is dedicated to the search for very high-energy cosmic neutrinos (GeV-PeV). KM3NeT/ORCA detector is optimized for the study of neutrinos created by cosmic rays in the Earth's atmosphere (a few GeV). To evaluate the efficiency and overall performance of neutrino telescopes Monte Carlo simulations are employed.

Graph Neural Networks (GNNs) are a specific class of neural network architecture designed to operate on graph-structured data. Such networks have been also developed in the context of the KM3NeT experiment. The input data of the GNN consists of nodes that store the hit (photomultiplier signal) information (time, Time-over-Threshold, 3d position, 3d direction) and edges that connect the nodes defining the relations between hits. The model architecture adopted for the analysis is based on the ParticleNet architecture [1].

In this study, the performance of the Graph Neural Network (GNN) is investigated for event classification and energy prediction using KM3NeT/ARCA data collected from 6, 7, and 8 DUs. Three main networks have been trained and evaluated using KM3NeT/ARCA8 (KM3NeT/ARCA with 8 detection units deployed) data: signal/background classifier, track/shower classifier and energy regression. Also, a signal/background classification network has been trained and evaluated using KM3NeT/ARCA7 data (KM3NeT/ARCA with 7 detection units deployed). The classification models have been trained to distinguish between atmospheric muons and signal (neutrinos), and shower-like and track-like neutrinos, respectively. The regression model is defined to produce an estimation of the neutrino energy. Moreover the performance of the KM3NeT/ARCA6 trained network is evaluated for the signal/classification task, using KM3NeT/ARCA7 and KM3NeT/ARCA8 data.

Περίληψη

Τα νετρίνα είναι υποατομικά σωματίδια πολύ μικρής μάζας, είναι ηλεκτρικά ουδέτερα και αλληλεπιδρούν μόνο μέσω βαρυτικών και ασθενών αλληλεπιδράσεων. Η τροχιά τους δεν επηρεάζεται από μαγνητικά πεδία, εξασφαλίζοντας έτσι την άφιξή τους στη Γη από απομακρυσμένες πηγές. Η πιθανή παρατήρηση νετρίνων από αυτές τις πηγές αναμένεται να παρέχει πληροφορίες για τις διαδικασίες που λαμβάνουν χώρα στο εσωτερικό αστροφυσικών αντικειμένων καθώς καθίσταται εφικτή και η παρατήρηση των ίδιων των πηγών. Επιπλέον, αποδείξεις για μη-μηδενική μάζα διευκολύνουν τις μελέτες σχετικά με τις ταλαντώσεις νετρίνο ("neutrino oscillations"), συμβάλλοντας στην περαιτέρω κατανόηση της θεμελιώδους φυσικής των στοιχειωδών σωματιδίων.

Η ανίχνευση νετρίνων γίνεται μέσω ειδικά κατασκευασμένων ανιχνευτών, οι οποίοι στηρίζονται στο φαινόμενο Čerenkov. Τέτοιοι ανιχνευτές χρησιμοποιούνται στο πείραμα KM3NeT για τον εντοπισμό νετρίνων υψηλής ενέργειας, στο εύρος GeV-PeV, όπου ως μέσο ανίχνευσης αξιοποιείται το θαλασσινό νερό. Το τηλεσκόπιο νετρίνων KM3NeT/ARCA στοχεύει στην αναζήτηση πολύ υψηλής ενέργειας κοσμικών νετρίνων (GeV-PeV), ενώ το τηλεσκόπιο KM3NeT/ORCA χρησιμοποιείται για τη μελέτη των νετρίνων που δημιουργούνται από κοσμικές ακτίνες στην ατμόσφαιρα της Γης (μερικά GeV). Επίσης, με τις προσομοιώσεις Monte Carlo γεφυρώνεται το χάσμα μεταξύ θεωρητικών προσδοκιών και πειραματικών παρατηρήσεων καθώς προσομοιώνονται οι αλληλεπιδράσεις των σωματιδίων.

Τα Νευρωνικά Δίκτυα Γράφων (GNNs) είναι ένας τύπος νευρωνικών δικτύων που χρησιμοποιούν δεδομένα με μορφή γράφου. Τέτοια δίκτυα έχουν αναπτυχθεί και στο πλαίσιο του πειράματος KM3NeT. Τα δεδομένα μορφής γράφου που δέχεται το δίκτυο αποτελούνται από κόμβους (nodes), που περιέχουν πληροφορίες σχετικές με τα "σήματα" ("hits") του ανιχνευτή (χρόνος, θέση - κατεύθυνση στις τρεις διαστάσεις), και ακμές (edges) που συνδέουν τους κόμβους μεταξύ τους ορίζοντας σχέσεις μεταξύ των hits. Το αρχιτεκτονικό μοντέλο που χρησιμοποιείται στην παρούσα ανάλυση είναι βασισμένο στο ParticleNet architecture.

Στην παρούσα εργασία μελετάται η επίδοση του GNN για την αναγνώριση σωματιδίων και την πρόβλεψη της ενέργειάς τους. Τρία βασικά δίκτυα έχουν αναπτυχθεί: ένα δίκτυο κατηγοριοποίησης σήματος/υποβάθρου, το οποίο μαθαίνει να ξεχωρίζει τα νετρίνα (σήμα) από τα ατμοσφαιρικά μιόνια (υπόβαθρο), ένα δίκτυο κατηγοριοποίησης νετρίνων track/shower, που μαθαίνει να ξεχωρίζει τα νετρίνα με υπογραφή καταιονισμών από τα νετρίνα που οδηγουν σε τοπολογίες γραμμικής τροχιάς στον ανιχνευτή, και τέλος, ένα δίκτυο για την πρόβλεψη της ενέργειας των νετρίνων. Όλα τα δίκτυα δέχονται δεδομένα από τον ανιχνευτή KM3NeT/ARCA8, τα οποία συλλέχθηκαν από 8 ανιχνευτικές μονάδες (DUs), με τα οποία γίνεται η εκπαίδευση (train) και η επαλήθευση. Επιπροσθέτως, μελε-

τάται το δίκτυο του σήματος/υποβάθρου στο οποίο έχει γίνει η εκπαίδευση με δεδομένα από KM3NeT/ARCA6 και KM3NeT/ARCA7.

Contents

1
Neutrinos

In 1892, Becquerel made the discovery of the radioactive phenomena. The fact that the electron is released from the nucleus during beta decay was originally noticed by Bohr. Chadwick made the important finding that the fundamental beta spectrum is continuous in 1914. This conclusion was thought to have different interpretations until far into the 1920s. It was understood that the neutron, which Chadwick discovered in 1932, was the particle in the nucleus that released the electron and neutrino during beta decay. Using the neutron, electron, proton, and neutrino, Fermi created the four Fermi-Hamiltonians for beta decay. The discipline of weak interactions emerged as a new area of theoretical physics; initially introduced by Fermi in 1933, it underwent further refinement in the 1950s through the contributions of Lee and Yang, Feynman and Gell-Mann, and many others, and finally put into its present form by Glashow, Weinberg and Salam in the 1960s. As for the strong forces, after Yukawa's groundbreaking work in 1934, there was a lack of a comprehensive theory until the development of chromodynamics in the 1970s. The neutrino was discovered by Reines and Cowan in 1956. Their method of detection involved the simultaneous observation of a positron emitted in the reaction $\overline{\nu}_e + p \rightarrow e^+ + n$ and a photon emitted in the deexcitation of cadmium following the neutron capture: $n +^{112} Cd \rightarrow^{113} Cd + \gamma$. For this purpose, they used the nuclear reactor, located at Savannah River in South Carolina, as both an anti-neutrino source and a detector of water containing dissolved cadmium chloride. Investigating beta decays has played a crucial role in unraveling key aspects of the weak force, including the chirality of neutrinos (their left-handed nature), the V-A structure of weak interactions, and the observation of parity non-conservation effects. Weak interactions gained recognition as a new force of nature when an increasing amount of particles were found to engage in them, and neutrinos became an essential component of these interactions.

1.1. Standard Model

The "Standard Model" [2] of elementary particles and forces is a theoretical framework derived from observations, foreseeing and interconnecting new data. Its success in predicting various phenomena has been remarkable. While it is not anticipated for the standard model to hold true at extremely short distances, its extraordinary achievements imply that it will likely continue to be a highly accurate

Quarks: $S=\frac{1}{2}$				Leptons: $S=\frac{1}{2}$				Gauge bosons: $S=1$	
$Q=\frac{2}{3}$	m	$Q=-\frac{1}{3}$	m	$Q=-1$	m	$Q=0$	m	quanta	m
$u_1\ u_2\ u_3$	$(2-8)10^{-3}$	$d_1\ d_2\ d_3$	$(5-15)10^{-3}$	e	5.11×10^{-4}	ν_e	$<1.5\times10^{-8}$	$g_1\cdots g_8$	$<$ a few $\times10^{-3}$
$c_1\ c_2\ c_3$	1.0–1.6	$s_1\ s_2\ s_3$	0.1–0.3	μ	0.10566	ν_μ	$<1.7\times10^{-4}$	γ	$<6\times10^{-25}$
$t_1\ t_2\ t_3$	173.8 ± 5.0	$b_1\ b_2\ b_3$	4.1–4.5	τ	1.7770	ν_τ	$<1.8\times10^{-2}$	$W^{\pm},Z^0$	$80.39\pm0.06,91.187\pm0.002$

Figure 1.1: Elementary particles of the standard model, consisting of gauge bosons and the three generations of fermions. $S\hbar$ is spin, Q_e is electric charge, and $m\,(GeV/c^2)$ is mass. Image credit [2].

approximation of natural phenomena even at scales as minute as 10^{-18} meters. In the early 1960s, physicists characterized nature by four distinct forces, each with different ranges and strengths measured at an energy scale of around 1 GeV. The strong nuclear force operates over a range of approximately a fermi or 10^{-15} meters. The weak force, responsible for radioactive decay and operating within a range of 10^{-17} m, is about 10^{-5} times weaker at low energy. The electromagnetic force, governing much of macroscopic physics, has both limitless range and strength and is mediated by the photon. Gravity, the fourth force, also possesses an infinite range and a low-energy coupling, roughly 10^{-38}, making it too weak to be detected in laboratory experiments. The Standard Model's significant achievement was the development of a unified description encompassing the strong, weak, and electromagnetic forces within the framework of quantum gauge-field theories. Furthermore, the standard model integrates the weak and electromagnetic forces into a single electroweak gauge theory.

Leptons, quarks, and mediators are the three types of elementary particles that make up all matter. Based on their charge (Q), electron number (L_e), muon number (L_μ), and tau number (L_τ), there are six different types of leptons. They divide into three generations by nature. Additionally, there are six corresponding antileptons with reversed signs. Thus, the total number of leptons is really 12. Quarks may also be categorized into six "flavours" based on charge, strangeness (S), charm (C), beauty (B), and truth (T). In addition, there are downess (D) and upness (U), even though these are not very common terms. Quarks are also classified into three generations. All signs would also be inverted on the antiquark table. Furthermore, there are three colours associated with each quark and antiquark, hence adding up to 36 quarks. All fermions, leptons and quarks, have a half-unit intrinsic angular momentum. The first generation comprises the up (u) and down (d) quarks that make up nucleons, along with pions and other mesons involved in nuclear binding. It also contains the electron e and electron-neutrino ν_e released by nuclear β decay with a positron. The quarks of the other two generations, charm (c) - strange (s) and top (t) - bottom (b), are constituents of heavier, short-lived particles. They and their accompanying charged leptons, muon μ and tau τ, decay swiftly via the weak interaction to the first family's quarks and leptons.

Lastly, each interaction has a corresponding mediator: the photon for electromagnetic force, two W's and a Z for weak force, the graviton for gravity, and the gluon for strong force. In the standard model, there are eight gluons, each

possessing a color. Similar to quarks, gluons are not expected to exist as isolated particles. An overview of the elementary particles of the standard model can be seen in Figure 1.1.

It is crucial to note that, within the standard model, neutrinos are considered to be without mass. This aspect will be explored in more detail in section 1.4.

All of this adds up to a significant quantity of elementary particles. Furthermore, there are at least 61 particles to deal with because the Glashow-Weinberg-Salam theory requires at least one Higgs particle. Guided by our understanding, initially with atoms and later with hadrons, numerous individuals have proposed that at least some of these 61 entities may be compositions of more fundamental subparticles. However, such hypotheses extend beyond the standard model. One may wonder why there should be three generations instead of only one. In response, the dominance of matter over antimatter allows for a reasonable Standard Model explanation, but only in the case of at least three generations. Naturally, this raises the opposite query, which is: why there are only three generations. There were valid reasons to expect a fourth or perhaps a fifth generation in 1988. However, in less than a year, SLAC and CERN experiments closed off that possibility. The Z^o particle, known as the 'mother of all particles', can undergo decay into any quark/antiquark or lepton/antilepton pair, given that the particle's mass is less than half that of the Z^o. This ensures there is enough energy to produce the pair. By gauging the Z^o lifetime, it becomes possible to count the number of quarks and leptons with masses below 45 GeV/c^2. Experimental results align with expectations based on the established three generations.

In the meantime, several theoretical hypotheses, lacking direct experimental support, extend beyond the Standard Model. Grand Unified Theories (GUTs), connecting strong, electromagnetic, and weak interactions, are widely acknowledged. The concept of 'supersymmetry' (SUSY), doubling particle numbers by associating fermions with bosons and vice versa, is also appealing to theorists. Additionally, since 1984, superstring theory has captivated the imagination of a generation of particle theorists. Superstrings not only promise to reconcile quantum mechanics and general relativity, eliminating infinities in quantum field theory but also propose a unified 'theory of everything', wherein elementary particle physics, including gravity, emerges as an inevitable consequence.

1.2. Fundamental properties of neutrinos

In the 1920s, research on nuclear beta decay yielded the first indications of the existence of neutrinos. Beta decay is a type of radioactive decay in which a beta particle (either an electron, β^-, or a positron, β^+) is emitted from an atomic nucleus (A, Z) as well as a lighter secondary nucleus with the same mass number but differing by one unit, and nothing else observable:

$$(A, Z) \rightarrow (A, Z \pm 1) + e^{\mp} + nothing\ else\ visible. \tag{1.1}$$

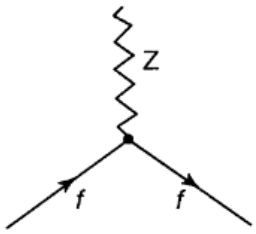
Experimentalists observed a continuum spectrum extending from m_e to the maximum permitted energy Q when the expected energy $E_e \sim Q = M_i - M_f$ for $e^{\pm}$ should have been well defined in the absence of any additional unseen particles in the final state. In 1930, Wolfgang Pauli proposed that during the beta decay process, an unobserved particle was emitted. The non-visible particle has to be neutral to conserve energy, extremely light, and with spin 1/2 to meet the requirements of statistics and angular momentum conservation. Beta decays are currently regarded as the following processes in contemporary notation:

$$\begin{aligned} &n \rightarrow p + e^- + \overline{\nu}_e, \\ &A(Z,N) \rightarrow A(Z+1, N-1) + e^- + \overline{\nu}_e, \\ &A(Z,N) \rightarrow A(Z-1, N+1) + e^+ + \nu_e. \end{aligned} \tag{1.2}$$

Three types of neutrinos have been identified via the development of new accelerators and advanced detectors. There is a unique family of leptons associated with each neutrino. Neutrino observations have been essential in helping us comprehend weak interactions, resulting in the standard model. Further observations using neutrinos are believed to provide an important contribution to our knowledge of physics beyond the standard model. In recent decades, it has become clear that learning about the neutrino is essential in our comprehension of the Universe. This is due to the fact that the neutrino is the most prevalent type of matter in the Universe after radiation, and it is the source of the heavy elements that are essential to life because of its role in nucleosynthesis. Thus, it is evident that the neutrino is a significant particle.

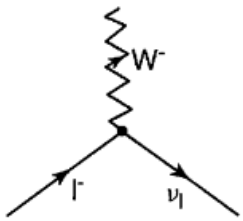
Neutrinos are several orders of magnitude lighter than all other fermions, and no direct measurement has found evidence for a non-zero mass. Neutrinos are also neutral; they do not perceive strong interactions and interact relatively weakly, if at all. For example, neutrinos produced in nuclear reactors with energy $E_\nu \sim 1\,MeV$ have a 10^{-11} chance of interacting within the Earth while traveling along a track that goes through its center. Neutrinos are therefore "elusive" particles, but they are also abundant in the Universe, yet we do not see them since they interact so infrequently. Neutrinos have two types of interactions:

- Neutral Current (NC), in which they couple with a Z^o boson, altering their 4-momentum but retaining their identity, and remaining uncharged. The fundamental neutral vertex looks like:



where f can be any lepton or quark. The Z mediates processes such as neutrino-electron scattering $\nu_\mu + e^- \rightarrow \nu_\mu + e^-$ and neutrino-proton scattering $\nu_\mu + p \rightarrow \nu_\mu + p$. In atomic physics, distinguishing neutral weak effects from electromagnetic processes is sometimes possible by exploiting the fact that weak interactions violate the conservation of parity (mirror symmetry). To observe a purely neutral weak interaction, researchers must turn to neutrino scattering, where there is no competing electromagnetic mechanism. The weak neutral currents were first observed in neutrino scattering $\nu_\mu e$ experiments at CERN in 1973, and the confirmation of their existence came shortly thereafter through experiments at Fermilab.

- Charged Current (CC), in which they couple with a $W^\pm$ boson, "transforming" into one of the charged leptons $e^\pm$, $\mu^\pm$ or $\tau^\pm$. The fundamental charged vertex looks like:



demonstrating the reaction $l^- \rightarrow \nu_l + W^-$. Charged weak interactions are the only ones that change the flavor of the neutrino. Such processes are the neutrino-muon scattering $\mu^- + \nu_e \rightarrow e^- + \nu_\mu$, the decay of the muon $\mu^- \rightarrow e^- + \nu_\mu + \overline{\nu}_e$, the decay of the pion $\pi^- \rightarrow e^- + \overline{\nu}_e$ and the beta decay of the neutron $n \rightarrow p^+ + e^- + \overline{\nu}_e$.

The flavor of a neutrino is really the type (that is the mass) of the charged lepton that is attached to the same charged current vertex. For instance, in the decay of a (virtual) W^- or W^+, one writes:

$$W^+ \rightarrow l^+ + \nu_l, \qquad W^- \rightarrow l^- + \overline{\nu}_l, \tag{1.3}$$

where the mass (or type) of the charged lepton is indicated by the label $l = e, \mu, \tau$ of the neutrino. There is a clear experimental explanation for the distinction between ν and $\overline{\nu}$ as well as the l label assigned to the neutrino. A "ν_e," as described by convention, interacting at its creation point can only produce e^-, whilst a "$\overline{\nu}_e$" can only produce e^+, a ν_μ only a μ^-, and similarly for the other neutrino types. Overall, there is a clear phenomenological significance to both the flavour label and the "bar" notation. The inherent angular momentum (or spin) of neutrinos is an integer. The remarkable discovery of the 1950s was that, given experimental uncertainties, the spin of every "ν" is anti-parallel to its momentum, yet the spin of every "$\overline{\nu}$" is parallel. Put otherwise, all of the $\overline{\nu}$ are right-handed and the ν are left-handed. This is the basis of the weak interactions' "chiral nature" and gives rise to the consequences of parity violations seen in nature.

1.3. Neutrino flavors

As aforementioned, there are three types of neutrinos in the particle physics standard model. Each neutrino completes a doublet with the corresponding charged leptons. A W^+ gauge boson, for instance, can couple to any of the three pairings of charged leptons and neutrinos:

$$\begin{aligned} W^+ &\to e^+ \nu_e \\ &\to \mu^+ \nu_\mu \\ &\to \tau^+ \nu_\tau \end{aligned} \tag{1.4}$$

The label given to the neutrinos corresponds to the flavour, which is the mass of the associated charged lepton.

Muon neutrino
Muon neutrinos and anti-neutrinos result from the decay of pions and muons, as illustrated in the reactions:

$$\pi^+ \to \mu^+ + \nu_\mu, \quad \mu^+ \to e^+ + \nu_e + \overline{\nu_\mu}. \tag{1.5}$$

The discovery of the decay $\pi^\pm \to \mu^\pm + \overline{\nu}_\mu$ raised the question of whether the undetected particles ν were the same as those produced in beta decays. In 1962, an experiment was conducted by L. Lederman, M. Schwarz, and J. Steinberg using accelerator neutrinos. By bombarding a Beryllium target with a 15 GeV primary proton beam, they generated pions, a source of neutrinos through decay. Studying interactions like $\nu_\mu + N \to \mu^- + X$ and $\nu_e + N \to e^- + X$, the experiment found only the first type, demonstrating the distinction between ν_μ and ν_e.

Electron neutrino
As mentioned above the electron neutrinos and anti-neutrinos are produced in beta decay processes $n \to p^+ + e^- + \overline{\nu}_e$.

Tau neutrino
The tau neutrino ν_τ is linked to the tau τ lepton, which is the third and heaviest of the charged leptons in the Standard Model of particle physics. The tau lepton τ^- was discovered in 1975 at the Stanford Linear Accelerator Center (SLAC) in California. In processes involving the decay of tau particles $\tau^\pm$, tau neutrinos ν_τ and anti-neutrinos $\overline{\nu_\tau}$ are expected to be produced as,

$$\begin{aligned} \tau^- &\to \nu_\tau + W^-, \\ &\to \nu_\tau + e^- + \bar{\nu}_e, \\ &\to \nu_\tau + \mu^- + \bar{\nu}_\mu, \\ &\to \nu_\tau + [\bar{u}s']. \end{aligned} \tag{1.6}$$

Measurements of the Z-boson width have ruled out a fourth neutrino if it is less than 40 GeV. If a fourth form of neutrino exists, it would indicate the existence of a fourth type of matter, which would be a significant finding.

1.4. Neutrino oscillations

The probability of detecting a neutrino, initially created in a specific flavor state, in the same state or transitioning to another flavor state can exhibit oscillations over time [3]. The concept of neutrino oscillations was first introduced by Bruno Pontecorvo in the 1950s, suggesting that if neutrinos possess mass, they could undergo such flavor-changing oscillations. This intriguing phenomenon stems from the principles of basic Quantum Mechanics. Neutrinos produced in charged current weak interactions alongside a charged lepton exist as weak eigenstates ν_e, ν_μ, ν_τ. Typically, these weak eigenstates lack a well-defined mass and can be expressed as linear combinations of three states, namely ν_1, ν_2, and ν_3, each characterized by distinct masses (m_1, m_2 and m_3). The expression for neutrino oscillations in vacuum using the Dirac formalism is:

$$|\nu_\alpha\rangle = \sum_j U^*_{\alpha j}|\nu_j\rangle, \tag{1.7}$$

where $|\nu_\alpha\rangle$ denotes the eigenstates ν_e, ν_μ, ν_τ, $|\nu_j\rangle$ represents the states ν_1, ν_2, ν_3, and $U^*_{\alpha j}$ is the unitary Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix. The unitary arises due to the orthogonality of the three flavor eigenstates and masses. The time evolution of a mass eigenstate with well-defined energy involves a simple phase factor, and for a state initially created with a well-defined flavor, the time evolution results in

$$|\nu(t)\rangle = \sum_j U^*_{\alpha j} e^{-iE_j t}|\nu_j\rangle. \tag{1.8}$$

The oscillation probability is

$$P(\nu_\alpha \to \nu_\beta; t) = |U_{\beta j} U^*_{\alpha j} e^{-iE_j t}|^2. \tag{1.9}$$

This probability exhibits periodicity in time, and notably, it does not vanish, even for $\beta \neq \alpha$.

If the neutrinos pass through a material medium their oscillation behavior is modified. This phenomenon is known as the Mikheyev-Smirnov-Wolfenstein (MSW) effect [4]. The existence of matter adds an extra component to the interaction for electron neutrinos since ordinary matter contains electrons but lacks muons or taus. Consequently, when a ν_e beam traverses through matter, it undergoes both charged and neutral interactions with electrons. On the other hand, ν_μ and ν_τ exclusively engage in neutral current interactions, resulting in a difference in the magnitude of their interactions compared to ν_e. These interactions cause alterations in the effective mass that a particle exhibits while traveling through a medium. As a consequence, oscillation probabilities deviate from their vacuum values.

An experiment that seeks to detect atmospheric neutrinos in order to determine the parameters that regulate flavor transitions is KM3NeT/ORCA (see Chapter 2).

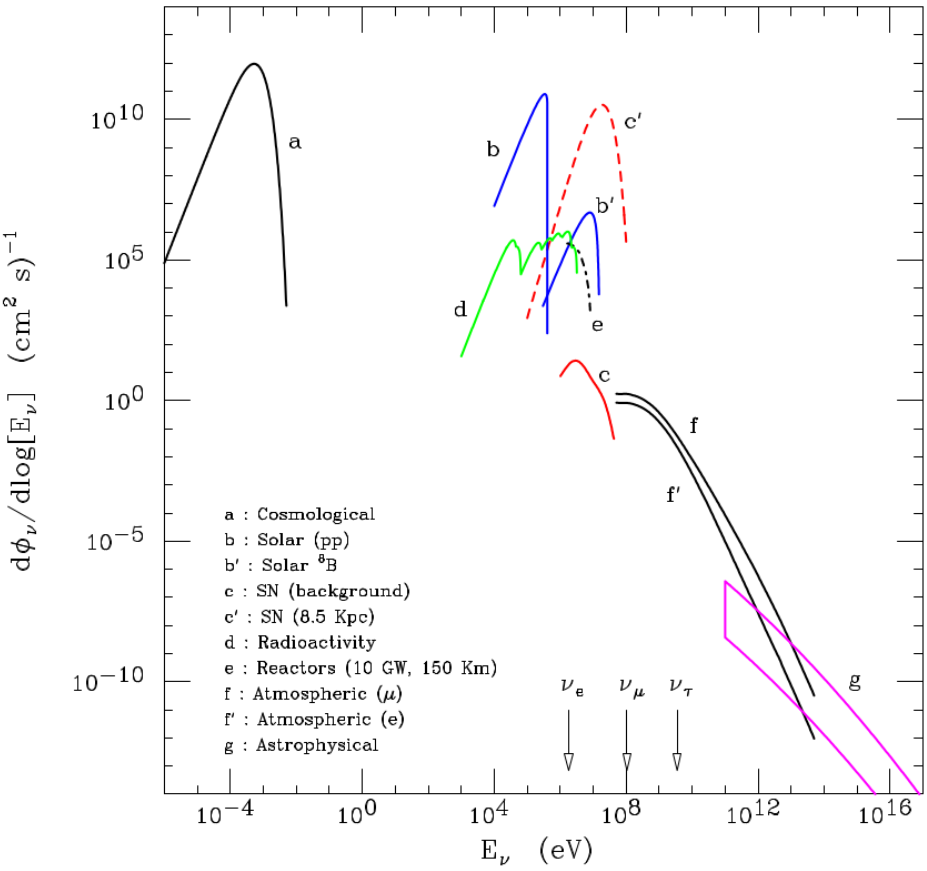


Figure 1.2: The flux of neutrinos on the surface of the Earth. The energy thresholds for charged current interactions on a free proton target are illustrated by three arrows. The line representing cosmological neutrinos assumes a vanishing neutrino mass. The line corresponding to Supernovae neutrinos describes only $\overline{\nu}_e$. The calculation of atmospheric neutrino fluxes is based on the Kamioka location, with only the lowest energy segment depending on the location. Image taken from [3].

1.5. Neutrino sources

Neutrinos are ubiquitous particles in the cosmos and hold significant importance in various astrophysical and cosmological phenomena. Figure 1.2 illustrates a graph depicting the energy distribution of neutrinos reaching the Earth's surface. This spectrum spans over 20 orders of magnitude, encompassing a wide range in both energy and intensity.

Cosmological neutrinos

Cosmological neutrinos [3] are neutrinos generated in the early stages of the Universe, with an approximate number density of 56 cm^{-3} for each neutrino species ($\nu_e, \overline{\nu}_e, \nu_\mu, \overline{\nu}_\mu, \nu_\tau, \overline{\nu}_\tau$) and a black-body spectrum at a temperature around 1.947 Kelvins. During the Universe hot phase, neutrinos were in equilibrium with photons. As the temperature dropped below 10^{10} Kelvin, neutrinos decoupled from the rest of the Universe but continued to cool with the Universe expansion. The neutrino

temperature is cooler than photons due to the reheating of photons from electron-positron annihilation. Current measurements of the Universe age and expansion rate provide strict constraints on neutrino masses. If neutrinos have masses on the order of a few eV or more, they could be a significant or dominant component, addressing the "Dark Matter Problem". Neutrinos, if crucial to the Universe mass, play a crucial role in forming observed structures like galaxies and galaxy clusters. Early on, the Universe was smooth, as indicated by small amplitude of temperature fluctuations in the cosmic microwave background radiation. Gravity enhances density contrasts, leading to the formation of structures. Neutrinos, being weakly interacting and remaining relativistic due to their small mass, efficiently erase initial fluctuations up to large scales. This contrasts with other forms of dark matter, such as WIMPs (Weakly Interacting Massive Particles), which move more slowly due to their larger mass and are less efficient in erasing fluctuations.

Solar neutrinos

Neutrinos [3] are abundantly produced through thermonuclear reactions occurring within stellar interiors, specifically in our Sun. A pivotal fusion reaction releasing nuclear binding energy is described as

$$4p + 2e^{-} \rightarrow^{4} He + 2\nu_e,$$

liberating 26.73 MeV. The considerable luminosity of the Sun suggests a substantial flux of ν_e reaching Earth. The nuanced energy spectrum of solar neutrinos is contingent upon the intricate nuclear reactions orchestrating the conversion of hydrogen to helium. While the majority of solar neutrinos possess energies below 0.41 MeV, a smaller yet significant component originating from the beta decay of Boron-8 extends up to 14 MeV, playing a crucial role in solar neutrino detection. The energy released in fusion reactions contributes to the Sun's thermal energy, with neutrinos successfully escaping and carrying away a proportion of the liberated energy. The solar neutrino flux is given by the equation

$$\Phi_{\nu_e} \simeq \frac{1}{4\pi d_{\odot}^2} \frac{2L_{\odot}}{(Q - \langle E_{\nu} \rangle)} \tag{1.10}$$

which is intricately linked to the solar luminosity $L_{\odot} \simeq 3.842{\times}10^{33}\ erg/s$, Earth-Sun distance $d \simeq 1.495x10^{13}\ cm$, and the average energy $E_{\nu} \simeq 0.3\,MeV$ conveyed by neutrinos in a fusion cycle. Despite variations in energy distributions arising from distinct reaction cycles, the predicted solar neutrino flux reaching Earth remains exceptionally high, estimated at approximately $6 \times 10^{10}\ cm^{-2}s^{-1}$. Comprehensive predictions necessitate the development of solar models that meticulously account for the contributions from various reaction cycles.

Supernovae

Type II Supernovae explosions [3] signify the culmination of the lifecycle of massive stars (with masses greater than the solar mass) that have evolved to develop an iron

core surrounded by successive burning shells and an outer envelope of hydrogen and helium. The iron core, being the most tightly bound nucleus in nature, undergoes collapse when it accumulates sufficient nuclear ash and reaches the Chandrasekhar limit of approximately 1.4 $M_{\odot}$. This collapse is rapid and results in a "neutronization burst" of electron neutrinos ν_e as the collapsing core converts nearly all protons into neutrons. The collapsing core's implosion is halted when it reaches nuclear density, leading to the formation of a shock wave that propagates outward, ejecting the outer layers of the star and causing a visible explosion. The newly formed proto-neutron star, with a radius of around 10 km, contains a significant amount of kinetic energy, primarily radiated away in the form of neutrinos. These neutrinos play a crucial role in the explosion, possibly depositing enough energy near the outward propagating shock to trigger the explosion.

All six neutrino flavors contribute roughly equally to the energy outflow, produced in the hot core through "flavor-blind" processes. The neutrino emission lasts for approximately 10 seconds, during which neutrinos undergo numerous scatterings in the dense material, determining the time of emission. The different average energies of neutrino components result from distinct cross sections for elastic scattering with electrons, with electron neutrinos ν_e having the largest cross section and are emitted from the regions closer to the surface of the stars, whereas neutrinos of the muon ν_μ and tau ν_{tau} flavors, possessing a smaller cross section, are emitted from deeper regions within the stars and are characterized as "hotter".

The theory of neutrino emission in supernovae explosions received confirmation on February 23, 1987, with the detection of neutrinos and radiation from supernova SN1987A, which had exploded 170,000 years earlier in the Large Magellanic Cloud. Detectors like Kamiokande and IMB observed a few events in coincidence with each other, providing information on the fluence and temperature of emitted neutrinos. These events also allowed for the determination of limits on neutrino properties, including their mass. The data showed no correlation between energy and time of arrival, leading to a conservative upper limit on the neutrino mass: $m_{\overline{\nu}_e} < 20\, eV$.

Radioactive neutrinos

The Earth emits about 40 Terawatts of energy, with 40% originating from the decay of radioactive nuclei, primarily from the Uranium and Thorium decay chains [3]. These chains involve specific decay cascades. In each β decay electron neutrinos $\overline{\nu}_e$ with a maximum energy of 3.27 MeV are emitted. The resulting geophysical $\overline{\nu}_e$ flux at the Earth's surface is estimated approximately at a few $10^6\,(cm^2\,s)^{-1}$ but it varies based on the uneven distribution of Uranium and Thorium in the crust and mantle, being depleted in the core. Detector location impacts the flux, with thicker crust regions, like Italy, having a larger flux than thinner crust areas, such as Japan. New large-mass detectors present an opportunity to observe these geophysical neutrinos, offering insights into Earth's structure and evolution.

AGN

Active galactic nuclei (AGN) [5] are powerful sources of high-energy neutrinos. AGN, fueled by the accretion of mass onto supermassive black holes, are not only the brightest sources of electromagnetic radiation but also potential accelerators of cosmic rays, including ultrahigh-energy cosmic rays. Neutrinos from AGN have been studied since the late 1970s, considering the interaction of accelerated protons with ambient photons. Early models suggested large diffuse neutrino intensities, but subsequent observations and refined models have constrained these predictions.

Atmospheric neutrinos

The Earth is exposed to a nearly uniform and consistent flow of cosmic rays, with an intensity of around 0.5 particles/(cm^2 sec sr). Atmospheric neutrinos [3] arise from the interactions of cosmic rays with the Earth's upper atmosphere. Cosmic rays are high-energy charged particles, primarily protons with contributions of completely ionized nuclei, and smaller components of electrons, anti-protons and positrons and originate from astrophysical sources. Their energy can extend up to 10^{20} eV. Galactic magnetic fields trap these cosmic rays for millions of years, causing their directions to become isotropic and temporally uniform. Upon reaching the upper atmosphere, primary cosmic rays collide with air nuclei, producing secondary particles such as protons, neutrons, muons and kaons. Neutrinos are prolifically generated in the ensuing cascades, with the primary source being the decay of charged pions $\pi^{\pm}$ and subsequent muon $\mu^{\pm}$ decays. The dominant decay channel involves π^+ decaying into μ^+ and a muon neutrino ν_μ, followed by μ^+ decay producing a positron e^+, an electron neutrino ν_e, and another muon neutrino ν_μ, such as:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu, \quad \mu^+ \rightarrow e^+ + \nu_e + \overline{\nu}_\mu$$

Similar processes occur in the decay of kaons.

1.6. Water Čerenkov detection

Experiments for astronomy through high-energy neutrino detection use Čerenkov detectors for their operation. These detectors derive their name from their fundamental operating principle, the detection of Čerenkov radiation. Čerenkov radiation is electromagnetic radiation emitted when a charged particle passes through a dielectric medium at a speed greater than the phase velocity (U_{ph}) of light in that medium. The emitted radiation is in the form of a characteristic cone of light, known as a Čerenkov cone. This emitted light forms a conical wavefront with a characteristic angle θ_C, determined by the medium refractive index n. An illustration of the Čerenkov cone and angle θ_C is shown in Figure 1.3. For particles with relativistic velocity $\beta = u/c$, the Čerenkov angle is

$$cos(\theta_C) = \frac{1}{\beta n}$$

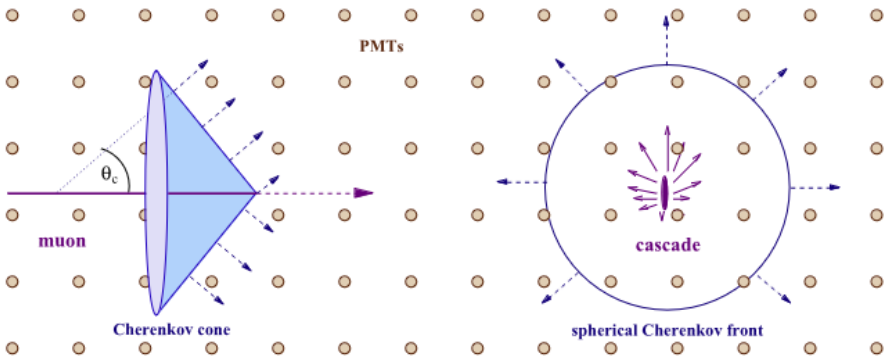


Figure 1.3: Illustration of the Čerenkov cone and angle θ_C. Detection principles for muon tracks (left) and cascades (right) in underwater/ice detectors. From [3].

and the refractive index of the medium is defined as $n = \frac{U_{ph}}{c}$. The condition for Čerenkov radiation to occur is $\beta n > 1$. Since the refractive index is associated with the frequency of radiation ($n = \frac{c}{f\lambda}$), there exists a cutoff frequency in the X-ray spectrum. The predominant radiation is in the ultraviolet, with visible blue radiation appearing for charged particles of high energy. In the case of relativistic particles ($\beta \simeq 1$) in seawater ($n \simeq 1.35$), the Čerenkov angle is approximately 42^o.

The neutrino interacts with the atomic electrons in hydrogen and oxygen atoms. Due to the MeV range of neutrino energies, the atomic binding energies are insignificant, allowing the scattering to be treated as elastic scattering of neutrinos off free electrons. Consequently, the electrons gain some energy from the neutrino. If the final electron possesses sufficient energy, its speed exceeds the speed of light in water, leading to the emission of Čerenkov radiation from the electron [6]. Photo sensors in the detector capture the emitted light, allowing the reconstruction of the interacting particle's properties from the hit distributions. Large-scale neutrino telescopes like KM3NeT use this technique, positioning photo sensors a few meters to tens of meters apart to cover volumes up to the km^3 scale. However, there is a lower energy threshold E_{th} for the production of Čerenkov radiation, where the rest mass of the particle (m_0) must allow it to move faster than $\beta > 1/n$. The threshold energy is

$$E_{th} = \frac{m_0}{\sqrt{1 - \frac{1}{n^2}}}.$$

For electrons, this threshold is $T_{th} \simeq 0.25\,MeV$, and for muons, it is $T_{th} \simeq 53\,MeV$, representing the minimum kinetic energies required ($T = E - m_0c^2$).

The interaction of neutrinos can result in two scenarios within the detector. In the case of a neutral current interaction, the outcome is solely a hadronic shower, and the neutrino goes undetected. On the other hand, for charged current interactions, a hadronic component is present along with a lepton whose flavor is determined by the incident neutrino's flavor. Due to differences in mass and stability among electrons, muons, and taus, the Čerenkov light emission signatures vary

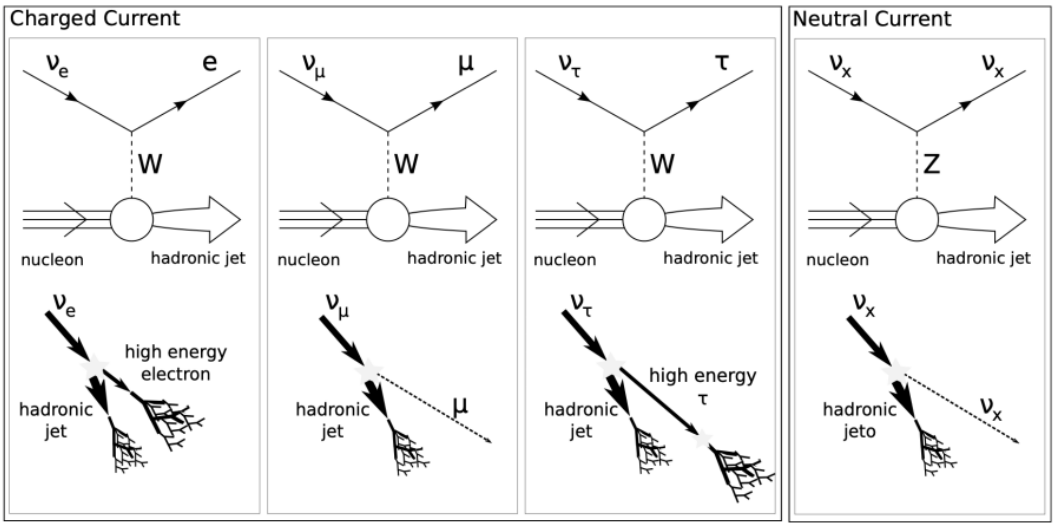


Figure 1.4: Signatures generated in the charged current and neutral current interactions for each neutrino flavor. From [7].

for each lepton flavor.

1.6.1. Shower/track signatures in water Čerenkov detectors

Using the light production profile's topology, which is ultimately the measured feature in Čerenkov detectors, we can categorize event signatures into two main types: showers and tracks. The contributions of each flavor to these classes are depicted in Figure 1.4. Particles with significant electromagnetic interactions that release energy primarily through ionization, called tracks, or through bremsstrahlung, called showers.

Hadronic and electromagnetic showers

Electromagnetic and hadronic cascades are phenomena that occur when high-energy particles interact with matter, leading to the production of secondary particles and the subsequent development of cascade-like events.

Electrons, with energies in the GeV scale, lose kinetic energy by emitting bremsstrahlung photons as they traverse the medium. These high-energy photons can undergo pair production, creating electron-positron pairs. Electrons and positrons generated in the cascade may scatter via Compton scattering or other processes. This cascading process forms an electromagnetic shower, where electrons and positrons with energies above the Čerenkov threshold (0.25 MeV) produce visible light. The overall signature of electromagnetic cascades appears point-like in detectors due to the 36 cm attenuation length of electrons in water. The number of Čerenkov photons is proportional to the cascade energy, enabling a precise energy reconstruction. Directional reconstruction is challenging due to the spherical propagation of light, but the emission peak aligns with the incident neutrino direction.

Hadronic cascades involve the production of additional hadrons, often light mesons, from the initial energy transfer to nucleons. The strong force governs the creation of hadrons, following the law of confinement. All participating hadrons with energies above the Čerenkov threshold emit light, with the overall light yield smaller than electromagnetic cascades due to higher particle masses and thresholds. Despite having a longer attenuation length (83 cm in water), the dimensions of hadronic showers are comparable to electromagnetic showers. Uncharged pions, the lightest mesons, are abundant and predominantly decay into two photons, contributing an electromagnetic component to the shower. While TeV and PeV showers look similar, lower-energy showers exhibit considerable event-by-event variation due to variations in the number, type, and propagation of final-state hadrons. In Čerenkov detectors only charged particles produce observable light, and the emission direction of hadrons deviates more significantly from the incident neutrino direction.

Muon tracks

Muons, whether generated directly in the atmosphere as secondary particles of cosmic rays or produced during interactions involving muon neutrinos, have distinctive characteristics that affect their behavior in water. Muons have the ability to travel significant distances in water, extending up to kilometers, depending on their energy. The primary process leading to energy loss in muons as they traverse through water is ionization, where interactions with atoms lead to the ionization of atoms along their path. For muons with energies exceeding 1 TeV, stochastic processes such as bremsstrahlung, pair production, and photonuclear interactions become more pronounced, causing significant energy losses and the creation of electromagnetic showers along the muon track. In the few GeV energy range, muons exhibit behavior similar to minimum ionizing particles, experiencing nearly constant energy loss over a straight path until they come to a stop or undergo decay. In a water medium, the length of a muon track is approximately 4 meters per GeV.

The shower topology includes neutral current (NC) interactions of all neutrino flavors and charged current (CC) interactions of ν_e and ν_τ , being the decay length of the τ lepton too short to be resolved below $\sim 1\,PeV$. On the other hand, events with track topology are produced by CC interactions of ν_μ.

2
KM3NeT neutrino telescope

KM3NeT [8] is a research infrastructure hosting two large-volume Čerenkov neutrino detectors which are currently under construction in the depth of the Mediterranean Sea. KM3NeT has two primary scientific objectives: the discovery and observation of high-energy neutrino sources in the Universe and the determination of the relative masses of the neutrinos, known as the neutrino mass hierarchy. To accomplish these goals two types of neutrino detectors (ARCA & ORCA) are established at different locations, both using the same technology but with distinct configurations.

The KM3NeT/ARCA detector (**A**stroparticle **R**esearch with **C**osmics in the **A**byss) is located approximately 100 km south-east of Sicily, Italy, off-shore the small town of Portopalo di Capo Passero in a depth of 3500 m. The ARCA neutrino telescope is optimized for the detection of high-energy neutrinos, in the GeV-PeV energy range, from astrophysical sources. Due to its advanced angular resolution capabilities, the investigation of the cosmic neutrino flux is enabled providing information about its source, energy spectrum and flavour composition.

The KM3Net/ORCA detector (**O**scillation **R**esearch with **C**osmics in the **A**byss) is located approximately 40 km off-shore Toulon, France, in a depth of 2450 m. The ORCA neutrino telescope is optimized for the detection of atmospheric neutrinos, in the few-GeV energy range, created in the Earth's atmosphere by cosmic rays. Exploiting the abundant fluxes of neutrinos, their fundamental properties can be studied, including the neutrino mass hierarchy and oscillation parameters.

A third location where the KM3NeT infrastructure will be implemented is off-shore Pylos, Greece, in a depth of 4550 m.

2.1. Technical description of the KM3NeT detectors

As mentioned above, both neutrino detector sites are using the same technology as well as the same detection mechanism, which is exploiting the Čerenkov emission produced by relativistic particles in neutrino interactions.

The infrastructure consists of 115 detection units (DUs) each of them giving vertical support for a string of 18 digital optical modules (DOMs). The DOM [9], [8] is a transparent 17-inch glass sphere comprised of two hemispheres, hosting 31

Figure 2.1: Digital optical module (DOM). Image credit [8].

Figure 2.2: Hamamatsu photomultiplier tube.

3-inch Hamamatsu photomultiplier tubes (PMTs). A DOM and a PMT are illustrated at figures 2.1 and 2.2, respectively. Inside a DOM the PMTs form five rings consisting of six PMTs each with an additional PMT positioned vertically downward at the bottom. At the lower hemisphere there are 19 PMTs and at the upper 12, held in place by a 3D printed support. The angle between each consecutive PMT in a ring is 60^o and each ring of PMTs is rotated by 30^o relative to the previous one. This geometric arrangement provides comprehensive coverage for detecting light signals from different directions. Each PMT is equipped with a light concentrator ring, enhancing the light collection area by 20% - 40%. The cavities between the glass and the support are filled with gel to ensure optical contact. Additionally, the DOM includes three calibration sensors: 1) an acoustic piezo sensor to determine its position, 2) compass and tilt meters to determine its orientation, 3) and LED nano-beacon for timing calibration. Inside the glass sphere are also housed the readout electronics. The specific components of the readout electronics are as follows: A Central Logic Board (CLB) is used to handle the digitized signals from all the sensors within the DOM. Two Octopus Boards connect groups of photomultiplier tubes (PMTs) to the CLB. Each Octopus Board links two groups of 19 + 12 PMTs to the central logic processing. There are 31 Base Boards in the DOM. These boards are responsible for amplifying and discriminating the signals from the PMTs. They also handle the high-voltage supply (1500 V) for the PMTs while having very low power consumption, just 45 mW. Finally, a Power Conversion Board (PCB) is in charge of supplying power to the entire DOM. Due to its structure the DOM can withstand the hydrostatic pressure encountered at the operating depths.

In Fig. 2.3 is illustrated the configuration of the detection units (DUs), also referred to as detection strings. Each string consists of two vertical, thin Dyneema® ropes, 4mm in diameter, attached to the DOMs via a titanium collar that surrounds the glass. Between the ropes are inserted additional spacers to maintain their parallel alignment. The backbone of the connections is obtained by a vertical electro-optical cable covered in plastic and filled with oil, balancing the pressure.

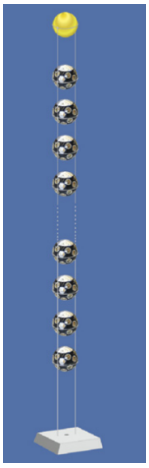


Figure 2.3: A schematic representation of a Detection Unit (DU). Image credit: [8].

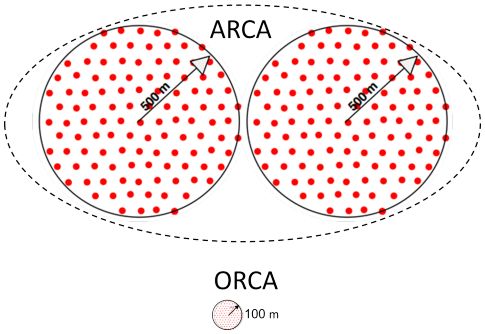


Figure 2.4: The respected sizes of the ARCA and ORCA neutrino telescopes. Image credit: [8].

Power transmission is facilitated along the DUs through a set of copper wires, while signals are conveyed to the shore via 18 optical fibers, one per DOM. A penetrator is employed to introduce the conductors and optical fibers into the glass sphere. To maintain the stability of the DU, a buoy is keeping the DU almost vertical while its base remains firmly anchored to the seabed.

The analogue signals from all the sensors in a DOM are digitized and then processed by the CLB, which is the core processing unit of the DOM. It captures the occurring time and the Time-over-Threshold (ToT) for each detected pulse with a precision of $1/\sqrt{12}ns$. The occurring time indicates the arrival time of the first photon, which is defined as the moment the voltage amplitude exceeds the threshold of 0.3 photo-electrons. The ToT corresponds to the length of time during which the amplitude of the signal remains above the threshold before it decreases below the threshold once again. The digitized data, that have passed the preset threshold, are sent onshore for further processing.

The KM3NeT/ARCA & ORCA detectors differ at their respective volumes and granularity. A comparison between those two is depicted in Fig. 2.4. In its completed state, KM3NeT/ARCA will comprise two building blocks of 115 vertical DUs spaced 90m apart, reaching a height of $700m$. The vertical space of the DOMs on the DUs is $36m$. Overall, it will cover an area with a radius of approximately $500m$ and have an instrumented volume of 1 cubic kilometer (km^3). On the other hand, KM3NeT/ORCA, once completed, will consist of a single building block with 115 DUs spaced $20m$ apart, resulting in a total height of $180m$. The vertical space of DOMs is $9m$. This configuration will cover an area with a radius of approximately $107m$ and have an instrumented volume of about 0.0067 km^3, providing new opportunities for neutrino oscillation studies in the Mediterranean.

2.2. Data acquisition

The KM3NeT detector utilizes a readout system [8] that involves sending all digital data from the photomultiplier tubes to shore if they exceed the predetermined threshold of 0.3 photo-electrons. Once on shore, the data is processed in real time on a computing cluster.

All data are sent from the detector to shore via the fibre-optic transmission system. The optical data includes the arrival time, the Time-over-Threshold (ToT) and the PMT identification number, commonly known as a "hit". Each hit consists of 6 bytes of data, with 1 byte for the PMT address, 4 bytes for the time information, and 1 byte for the ToT. Each DOM has a data rate of 9-12 Mb/s in which contributes neutrino-induced events as well as some background sources. The background is dominated by decays of radio-active potassium (^{40}K) in water, bio-luminescence and atmospheric muons. The total data rate for a completed building block amounts to 25 Gb/s and the relative time offsets between any pair of DOMs are stable within 1 ns. Due to the vast amount of information it is required a reduction of the data rate and only the filtered data are stored on the disk. In order to preserve all the available information for the offline analyses, each event will contain a "snapshot" of all the data captured by the detector during the event.

In essence, the first stage of data processing is to collect the unfiltered data from DOMs and reorganise them into events using a software called DataQueue (DQ) [10]. Then, run the triggering algorithms through an Optical/Acoustic Data Filter (ODF/ADF) software. Finally, the processed data are stored into the disk using the DataWriters (DW) application.

2.2.1. Triggering

A series of specific filter algorithms implemented in software (DF) are utilized to separate the physics events from the background noise. Different levels of filters [8] are applied to the data:

- Level-zero filter (L0) is the initial filter applied to the analog pulses (unfiltered data). It sets a threshold for the pulses and is the only one carried out off-shore. L0 hits are also referred to as "snapshot hits" or simply "hits" [7].
- Level-one filter (L1) involves the coincidence of at least two L0 hits from different PMTs of one DOM within a fixed time window, typically 10 ns. L1 hits are referred to as "triggered hits".
- Level-two filter (L2) further refines the data by reducing random coincidences using the PMTs' orientations.

In the DataFilter (DF) processing stage, three primary types of data are generated as outputs [11]:

- Triggered Events (EVT or IO_EVT): These events contain a "snapshot". This snapshot includes the information of the triggered hits as well as the background hits recorded during a specific time interval before and after the triggered hits.
- Summaryslice Data (SUM or IO_SUM): A summary slice is generated and recorded for the entire detector every 0.1 seconds. It encapsulates essential information about the detector performance and status. Within each summary slice there are summary frames, each of them dedicated to one DOM and they hold information about the Module Identifier, which is a unique code assigned to each DOM to distinctly identify it, the Average Rate for each of the 31 PMTs within each DOM for every 0.1 seconds, and the Status Bits which convey specific information about the status of the detector during that interval.
- Timeslice Data (TS or IO_TS): A time slice is generated and saved for the entire detector at regular intervals of 0.1 seconds. Within this time slice, data is organized into frames. Each of these frames corresponds to a specific DOM and contains details about the hits that occurred on the respective DOM during the 0.1 seconds. A "hit" includes information about the raw (uncorrected) hit time, PMT channel, and Time Over Threshold.

The above three types of data form "runs" and each run has a duration of 6 hours.

2.3. Monte Carlo Simulation

Monte Carlo (MC) simulations have a multifaceted role in experimental physics. They aid in optimizing different components and configurations of the experiment, validating the authenticity of real experimental data through comparisons with modeling the interactions of particles and simulated results.

In KM3NeT, MC simulations have been employed to model and simulate various aspects of the detector behavior and the interactions of particles in it. They replicate how the particles interact with the medium (water) surrounding the detector and the following Čerenkov light production, and predict the detector response in terms of the signal detected by the PMTs. These simulations bridge the gap between theoretical expectations and experimental observations, enabling researchers to interpret the data collected by KM3NeT accurately and make meaningful contributions to the field of neutrino astrophysics.

2.3.1. Event generation

The initial phase of the simulation process involves generating particle fluxes. These incoming particles include neutrinos originating from astrophysical sources, as

well as atmospheric muons and atmospheric neutrinos. Moreover, the geometry of the detector and its environment is generated.

The simulation involves the propagation of neutrinos through the Earth, simulating interactions in rock and sea water, for the astrophysical and atmospheric fluxes of (anti-)neutrinos of all three flavors ($\nu_\mu, \overline{\nu}_\mu, \nu_e, \overline{\nu}_e, \nu_\tau, \overline{\nu}_\tau$). Both neutral-current (NC) and charged-current (CC) interactions are simulated. Neutrino events are generated using the program GENHEN.

Atmospheric muons constitute a significant and frequent background. To replicate this background accurately, the program MUPAGE is utilized, generating both single and multiple atmospheric muon events.

Background photons, originating from the decay of ^{40}K in the sea water and bioluminescence, are also simulated by adding random noise events at a rate of 5 kHz per PMT. Finally, the trigger algorithms are applied and the MC events that pass the triggering conditions remain and the on-shore triggering is simulated.

2.4. Background sources

The KM3NeT detectors can distinguish three distinct types of optical background sources that conceal neutrino-induced events: bioluminescence, radioactive decay processes in water, and atmospheric muons.

- *Bioluminescence* is the emission of visible light by living organisms through a chemical reaction. It is prevalent in the deep sea, where surface light is absent. This phenomenon is diverse, with larger animals like fish or jellyfish producing bursts of bioluminescence lasting seconds, while bioluminescent bacteria serve as a weaker yet continuous light source, contributing to the background rate increasing the chances of pure noise events firing a trigger. Factors such as sea current and seasonal changes affect the bioluminescent activity. To mitigate its impact, the high-rate veto (HRV) is employed in data acquisition systems. It monitors the hit count of a single PMT, excluding it from triggering if hits exceed 2000 per 100 μs.

- The *radioactive decays* in water refer to processes where unstable isotopes within the water undergo radioactive decay, emitting particles such as alpha, beta, or gamma rays. In particular, seawater contains a small fraction of about 0.04% of radioactive Potassium-40 (^{40}K). ^{40}K primarily undergoes two decay channels; a β decay, in which a ^{40}K nucleus transforms into a calcium-40 (^{40}Ca) nucleus by emitting a β particle and an antineutrino:

  $$^{40}K \rightarrow^{40} Ca + e^- + \overline{\nu}_e,$$

  and an electron capture, where a ^{40}K nucleus captures one of its inner orbital electrons and converts a proton into a neutron. This process results in the

formation of argon-40 (^{40}Ar) emitting a photon in the subsequent gamma decay:

$$^{40}K + e^- \rightarrow ^{40} Ar^* + \nu_e \rightarrow ^{40} Ar + \gamma + \nu_e.$$

The electron, with a maximum energy of 1.31 MeV, has the capability to generate Čerenkov radiation. Similarly, the excited state of ^{40}Ar possesses an energy of 1.46 MeV, and its gamma ray can undergo scattering through the Compton effect, resulting in detectable photons emitted by the subsequent electrons. These decay processes of ^{40}K contribute to a consistent, uniform optical background within the detector. This background has the potential to trigger random L1 events, but it can also be intentionally employed for calibration purposes.

- *Atmospheric muons* are secondary particles produced in the Earth's atmosphere due to interactions between cosmic rays and air molecules. As they travel through air, water, or rock, emit Čerenkov radiation and lose energy along their path. As atmospheric muons lose energy due to Čerenkov radiation, they may eventually come to a complete stop. To minimize the impact of atmospheric muons as a background source in experiments, detectors are positioned deep below the Earth's surface. The overburden of rock, sea water or other materials above the detector acts as a shield, reducing the flux of atmospheric muons reaching the detector, yet they can penetrate to the detector volume if their energy at the sea surface is in the TeV range or above.

  For the KM3NeT/ARCA, the real-time trigger rate is dominated by downgoing atmospheric muons. Thus, trigger settings have been adjusted to maintain a manageable data rate of muons. To mitigate backgrounds from atmospheric muons and random coincidences of hits from ^{40}K decays, selection cuts are applied based on event reconstruction quality, reconstructed zenith angle for track-like events, and event energy-related quantities (e.g., number of hits) or event topologies.

2.5. Event reconstruction in KM3NeT/ARCA

Reconstruction involves adjusting model parameters to align with observed data. Relativistic charged particles emit Čerenkov radiation, forming a cone of light in our detector (see Section 1.6). Neutrino interactions are categorized as either *track* events, where a muon emerges, or *shower* events, generating cascades of particles. Different reconstruction algorithms are applied to handle each event type separately [12]. The underlying model assumes that, for both muons and particles in a shower, Čerenkov radiation is emitted as they traverse the detector.

2.5.1. Track reconstruction

Due to the non-linear nature of the task of fitting a muon track to PMT hit data, an approach with several consecutive steps is employed, referred to as the *JMuonChain* [13]. Describing the muon trajectory involves five distinct parameters: its direction, along with the time and position at various points along the trajectory. The muon trajectory fit employs a coordinate system, where the muon travels parallel to the z-axis and intersects the z = 0 plane at x_0, y_0 at time t_0. Assuming no scattering or dispersion of light, the expected arrival time t_i of Čerenkov photons on a PMT is defined as

$$t_i = t_0 + \frac{z_i}{c} + tan(\theta_C)\frac{R_i}{c},$$

incorporating the characteristic Čerenkov angle θ_C, the speed of light in a vacuum c, the minimum distance of the muon from the PMT $R_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}$, and the distance z_i from the PMT to the z = 0 plane.

A linear fit is used to address outliers from optical background hits and scattered photons in the data. The fit is applied to clusters of causally related hits, selecting the solution with the maximum number of hits unless a more statistically significant solution is found. Outliers are removed based on their contribution to the total χ^2. This process is repeated for various assumed track directions, where the N best-fit solutions are stored and used in the subsequent fit stage. The fit quality Q is quantified using $Q = NDF - 0.25 \times \left(\frac{\chi^2}{NDF}\right)$, where NDF is the number of degrees of freedom.

Starting from the twelve best-fit directions, a maximum-likelihood search is performed using the Levenberg-Marquardt method. The likelihood function incorporates probability density functions (PDFs) describing the PMT response:

$$\mathcal{L} = \prod_{PMThits} \frac{\partial P}{\partial t}(R_i, \theta_i, \phi_i, \Delta t), \tag{2.1}$$

where R_i is the minimum distance of the muon from the PMT, θ_i and ϕ_i describe the orientation of the PMT, and Δt is the time difference between expected and measured arrival of light. The PDFs include various information such as Čerenkov radiation, energy losses, dispersion, optical background, quantum efficiency, angular acceptance, and transit time spread of the PMTs. Once the muon trajectory is determined, the energy of the muon is fitted using the spatial distribution of hit and non-hit PMTs.

2.5.2. Shower reconstruction

The KM3NeT cascade reconstruction algorithm consists of two stages [12].

Initially, the shower vertex is fitted based on hit times, utilizing a high-purity sample of coincident hits within 20 ns on the same DOM. This sample may still contain background hits from optical sources, but their impact is mitigated by employing the M-estimator score-function $M = \sum_{hits} \sqrt{1 + (t_i - \hat{t}_i)^2}$, where t_i is the

hit time on the DOM i at a distance r_i, and $\hat{t}_i = t_0 + \frac{n}{c} r_i$ is the expected hit time of a spherical wavefront expanding with velocity c/n from the shower position at time t_0. The best fit for the shower position and time is determined using root's TMath::Minimizer interface [14] , starting from the center of gravity of the hits.

Subsequently, the shower direction and energy are estimated based on the distribution of hit or non-hit PMTs. The log-likelihood function used is purely based on hit/not hit information:

$$\mathcal{L} = \sum_{empties} log[P_i^{nohits}] + \sum_{PMThits} [1 - P_i^{nohits}]. \tag{2.2}$$

The probability is defined as $P_i^{nohits} = e^{-\mu_{sig}(r_i, z_i, a_i, E_S) - R_{bg} \cdot T}$, where only the expected number of photons $\mu_{sig}(r_i, z_i, a_i, E_S)$ on a PMT i in a time window T ranging from -100 ns to +900 ns are considered in order to minimize the optical background. E_S is the shower energy, and r_i, z_i, a_i the coordinates for the shower reconstruction as shown in Figure 2.5.

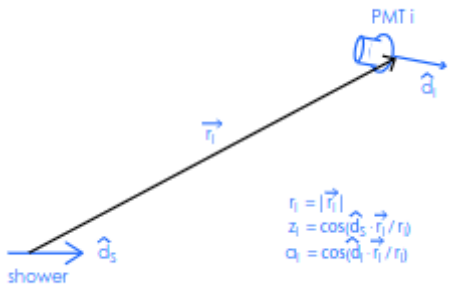


Figure 2.5: Coordinate system used in shower reconstruction. Image credit: [15]

2.6. Further neutrino experiments

The concept of using large volumes of natural material to detect neutrinos dates back to the mid-20th century, but the technical challenges of building and maintaining detectors in such environments were significant, and the necessary technologies were not fully developed at that time.

DUMAND [16] (**D**eep **U**nderwater **M**uon and **N**eutrino **D**etection project) was among the earliest concepts for high-energy neutrino and cosmic-ray detection using the deep ocean as a medium. The project was initiated in the 1970s and envisioned an array of detectors placed at 4800 meters deep underwater in the Pacific Ocean, off-shore Hawaii. It aimed at detecting high-energy particles by observing Čerenkov radiation. While facing technical and funding challenges, DUMAND played a crucial role in shaping the ideas and technologies that later led to successful experiments.

Building on DUMAND's concepts, the AMANDA [17] (**A**ntarctic **M**uon and **N**eutrino **D**etector **A**rray) experiment was developed as a precursor to IceCube. It was the first in-ice neutrino telescope located at South Pole. AMANDA deployed a network of optical modules deep within Antarctic ice to detect Čerenkov radiation produced by high-energy neutrinos interacting with the ice. Its construction started in 1993 and the detector was completed in 2000, operating with a total of 677 optical modules attached on 19 strings arranged in a circle with a diameter of 200 meters. The strings were deployed in a depth 1500 meters. On March 20, 2005, after nine years of operation, AMANDA and IceCube merge to form a single IceCube Collaboration.

2.6.1. Under-water detector: ANTARES

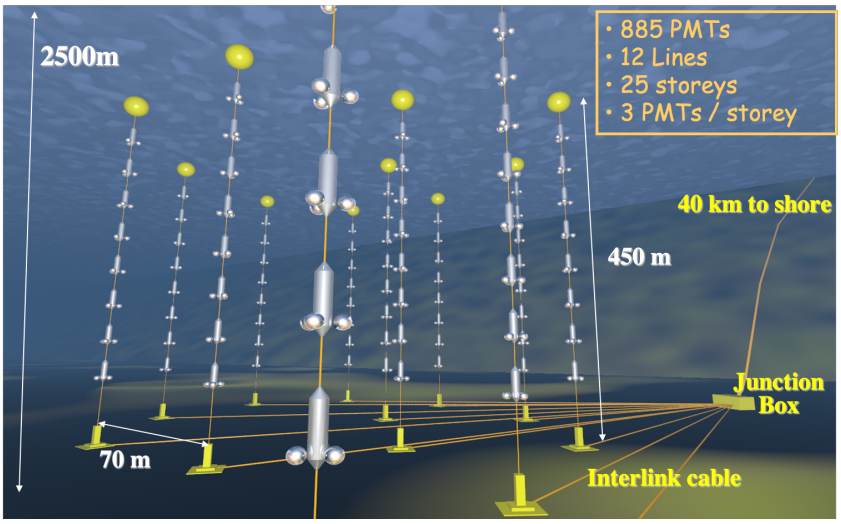


Figure 2.6: Schematic representation of the Antares neutrino telescope. Image credit: [15]

The ANTARES neutrino telescope [15] (**A**stronomy with a **N**eutrino **T**elescope and **A**byss environmental **RES**earch) is located approximately 40 km off-shore Toulon, France, in a depth of 2500 m, nearby the KM3NeT/ORCA site. It has an instrumented volume of 0.05 km^3, making it the largest observatory designed to detect neutrinos in the northern hemisphere and the first to operate underwater in the deep sea. The infrastructure consists of 12 lines of 300 storeys and 885 optical modules. A detector line is a vertical array of multiple storeys. Each detector line is formed by arranging 25 storeys with a vertical space of 14.5 meters. Each storey is enclosed by a titanium frame and comprises the optical modules as well as the necessary electronics required for data acquisition. An optical module is a pressure-resistant glass sphere that houses a single 10-inch PMT. Three such optical modules are attached on a storey with their PMTs having a 45^o deviation directed downward from the horizon. The schematic view of the ANTARES detector can be seen at Figure 2.6 and its components at Figures 2.7, 2.8. These detector lines are deployed vertically in the underwater environment where an anchor and a buoy

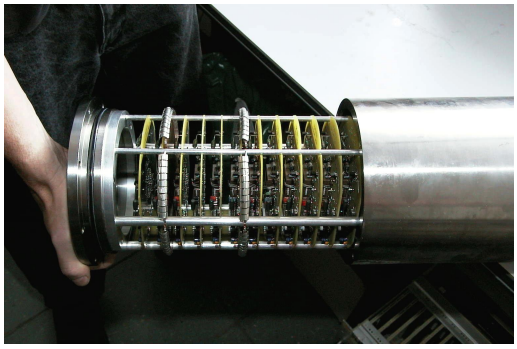


Figure 2.7: The local control module of a storey consisting the data acquisition electronics. Image credit: [15]

Figure 2.8: A storey with the three optical modules attached. Image credit: [15]

are keeping them in place. The first storey is placed at 100 meters above the seabed and the last one at about 450 meters.

The ANTARES scientific program aimes at advancing our understanding of the Universe through neutrino observations and it covered three main areas: Astronomy and Astrophysics, Dark Matter and Particle Physics, Neutrino Oscillations and Interactions. ANTARES primarily focuses on neutrinos of extraterrestrial origin, including those from astrophysical sources such as gamma-ray bursts, active galactic nuclei, microquasars and supernovae. Its location in the Mediterranean Sea complements the research conducted by other neutrino telescopes like IceCube and KM3NeT, providing a broader coverage of the sky.

The deployment of the first detector line occurred on February, 2006 and the ANTARES telescope was completed on 29 May, 2008. Nearly 16 years later, on February 12, 2022 the official cessation of data acquisition for the ANTARES detector was marked, initiating the process of its dismantling.

2.6.2. In-ice detector: IceCube

The IceCube neutrino observatory [18] is located near the Amundsen-Scott South Pole Station, Antarctica, burried below the surface in a depth of about 2500 meters. It is the first 1 cubic-kilometer particle detector made of 1 gigaton of Antarctic ice ever built. The layout of the detector is illustrated in Figure 2.11. The infrastructure consists of the surface array, IceTop, and a denser inner subdetector, DeepCore. The in-ice detector consists of 86 vertical strings, in and on each string are attached 60 DOMs frozen in boreholes, thus rising up to 5160 DOMs in total. The DOMs are placed in a 17 meter vertical distance from each other. The strings are arranged in a hexagonal grid pattern with a separation of 125 meters, covering an area of 1 km^3. The central part of the array features eight strings deployed more closely, with a horizontal separation of roughly 70 meters and a vertical spacing of 7 meters between DOMs. This more compact layout forms the DeepCore subdetector,

designed to reduce the threshold energy for neutrino detection to around 10 GeV. This configuration allows for the investigation of neutrino oscillations. The surfuce array, IceTop, is composed of 81 stations situated on the upper ends of the corresponding strings. Each of these stations includes two tanks, and within each tank, there are two downward-facing DOMs. This setup serves two main purposes. Firstly, it operates as a "veto and calibration detector" for the IceCube system and secondly, it is capable of detecting "air showers" initiated by high-energy primary cosmic rays. This process occurs in the energy range spanning from 300 TeV to 1 EeV.

IceCube's collaboration has achieved several significant milestones in neutrino astronomy. Among its notable achievements is the discovery of the diffuse flux of cosmic neutrinos as well as the detection of a 6.3 PeV neutrino through a Glashow resonance event. A significant expansion of the current IceCube neutrino telescope, *IceCube-Gen2* [19], is proposed. The plan involves extending the instrumentation to cover a volume of 10 km^3 within clear glacial ice at the South Pole. *IceCube-Gen2* primary objectives include identifying the sources of astrophysical neutrinos, discovering GZK (Greisen-Zatsepin-Kuz'min effect) [20],[21] neutrinos, and establishing itself as a key observatory in future multi-messenger astronomy initiatives.

Figure 2.9: The IceCube Laboratory. Image credit: IceCube Collaboration

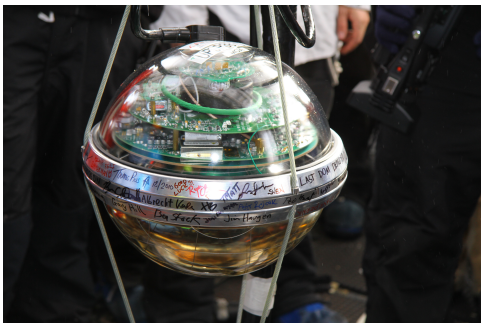


Figure 2.10: The last DOM before it was buried one mile deep in the Antarctic ice. Image credit: IceCube Collaboration

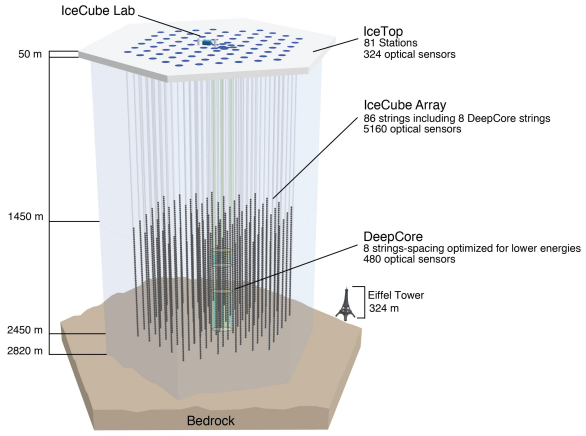


Figure 2.11: The IceCube Neutrino Observatory. Image credit: IceCube Collaboration [18]

3
Introduction to Machine Learning

Artificial Intelligence (AI) [22] is a rapidly evolving field with many practical applications and active research topics. AI focuses on developing systems and machines capable of simulating the human intelligence and even tackling and solving problems that are intellectually difficult for humans. The concept of AI emerged in 1950 when Alan Turing published a paper introducing the Turing test to determine whether a computer has the ability to exhibit human-like intelligence, but the term "Artificial Intelligence" was coined later, in 1957, by John McCarthy. In the early AI projects the knowledge base approach was used, where the systems relied only on hard-coded knowledge. Some difficulties on this approach led to the realisation that the systems should be capable of acquiring their own knowledge by extracting patterns from raw data. Hence, Machine Learning was introduced.

Machine learning (ML) is a subfield of artificial intelligence and computer science and focuses on building algorithms, extracting patterns from raw data and making accurate predictions on a given task in order to solve a problem. In general, machine learning framework involve the following three key components. The decision process, where the algorithms analyze the input data, which can be labeled or unlabeled, to generate estimations regarding patterns within the data. An error function is used to compare the model predictions to known examples, enabling the assessment of how well the model performs. And finally, the model is optimized by making adjustments to the model weights in order to minimize the discrepancy between the model predictions and the known examples. The performance of a simple machine learning algorithm is heavily affected by the representation of the data that it is given. The information included in the representation is known as "feature".

ML can be categorised into four main types: Supervised learning, Unsupervised learning, Semi-supervised learning and Reinforcement learning. A supervised machine learning algorithm is trained with a dataset containing labeled examples, whereas an unsupervised algorithm is trained on unlabeled data. Semi-supervised learning is a combination of both supervised and unsupervised learning, using a small labeled dataset to guide the training process and a larger unlabeled one for feature extraction. Reinforcement learning model is a model similar to the supervised one, yet it is not trained using labeled data, it learns to make sequential decisions by trial and error in predictions. For the purposes of this thesis the supervised method is used.

3.1. Deep Learning

Another part of AI is Deep Learning (DL) [22]. DL is an approach to machine learning that has seen tremendous growth in its popularity and usefulness, largely as the result of more powerful computers, larger datasets and techniques to train deeper networks. The main difference between Machine and Deep Learning is the way that each algorithm learns from data. Deep learning can learn from both labeled and unlabeled data and distinguish the important feature of the data. Thus, it can work directly with raw, unprocessed data, without the need for human experts to specify the features in advance. Contrarily, classical machine learning depends more on human guidance and the system needs to know in advance which features of the data to pay attention to. It typically works better with structured data, where the important features are well-defined.

Essentially, deep learning is a neural network with three or more layers. The term "deep" pertains to the depth of a neural network, composed of multiple hidden layers situated between the input and output layers. The initial layer, referred to as the "visible layer", is where the input data is introduced. It is called the "visible layer" because it contains the variables or features that we can directly observe in our data. The "hidden layers" are the subsequent layers in the neural network. They are referred to as "hidden" because they don't receive data directly from the input; instead, they are intermediate layers that extract increasingly abstract and complex features from the input data. These features are learned by the neural network during training. The neural network autonomously determines which features are useful for explaining the relationships and patterns within the observed data.

3.1.1. Structure of neural networks

A neural network or artificial neural network (ANN) consists of an input layer, hidden layers and an output layer. Each layer consists of *nodes*, also known as neurons (see Figure 3.1), where each node represents a feature or attribute of the input data. The *input layer* is responsible for receiving the raw data and passing it to the subsequent layers. Each *hidden layer* contains multiple neurons, and connections exist between neurons in adjacent layers. Each neuron in a hidden layer sums up the inputs it receives from its connections and applies an *activation function* f to produce its output. This output is then passed on to neurons in the next layer. The layers are connected by weighted connections, which determine the strength of the connections between neurons. Each connection has an associated *weight* w, which can be adjusted during training to optimize the network performance. Additionally, a *bias* term b is added to the sum of weighted inputs. For one neuron the output y is

$$y = f\Big(\sum_m w_m x_m + b\Big). \tag{3.1}$$

Activation functions are used to decide whether or not a specific neuron should

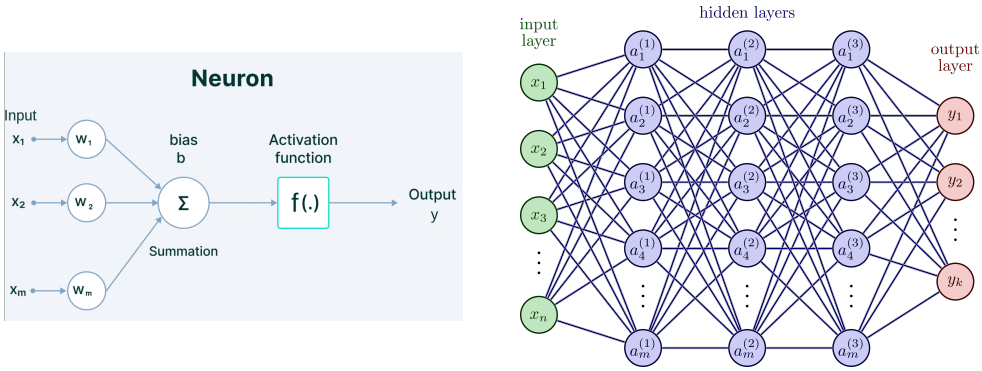


Figure 3.1: Left: Schematic drawing of a neuron with its components. Right: Structure representation of a neural network consisting of an input layer, three hidden layers and an output layer.

be activated. The choice of the function depends on the nature of the problem and the characteristics of the data. They can determine the neuron output based on this transformed input. There are two types of *activation functions*: linear and non-linear. However, linear functions are rarely used in deep networks due to their simplicity, making them unable to represent the complexity of the network. To introduce non-linearity and enable the network to learn complex relationships, a non-linear *activation function* f (e.g., ReLU, Sigmoid) is applied to the weighted sum of inputs and bias. The most common function due to its simplicity is the Rectified Linear Unit (ReLU) defined as,

$$f_{ReLU}(x) = \begin{cases} 0, & \text{if } x < 0. \\ x, & \text{otherwise.} \end{cases} \tag{3.2}$$

Typically, the ReLU *activation function* is preferred to be used at hidden layers and the *sigmoid activation function* at the output neuron which maps $\mathbb{R} \to (0, 1)$,

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{3.3}$$

An illustration of the ReLu and Sigmoid activation functions can be seen in Figure 3.2. If all outputs of one layer are connected to each input of the succeeding layer then the layers are called *fully-connected*. Those are the layers whose units receive as inputs the outputs of each of the units of the previous layer.

3.1.2. Training Neural Networks

Prior to commencing neural network training it is crucial to pre-process the input dataset. The first step is to shuffle the data to ensure that the they are not ordered in any specific way, thus preventing any unintentional patterns or biases in the data that may affect training. After shuffling, the input dataset is split into training and validation sets. The training set is used to train and build the model and is typically the largest dataset. The validation set is a smaller subset of the data. The model does not update the weights from this set. It is used to assess the performance of

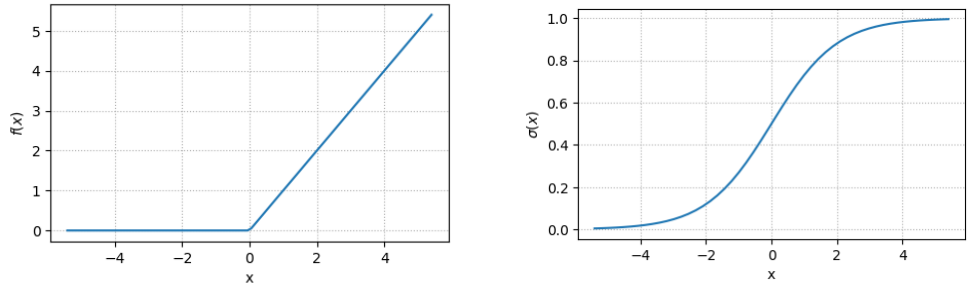


Figure 3.2: Left: Rectified Linear Unit (ReLU) function . Right: Sigmoid function, also known as logistic function.

the training in unseen data and helps to choose the learning algorithm and find the best values of hyperparameters. *Hyperparameters* are parameters that are set prior to the training and are not optimized during the learning process.

During the training process, input data are passed forward through the network. Each layer computes a weighted sum of inputs, applies an activation function, and passes the result to the next layer. For the network to be able to produce some meaningful predictions, the *weights* and *biases* of all connections are adjusted. A *loss function* is applied to the network predictions and the true values in order to minimize the loss.

Loss function

The *loss function* measures the difference between the network predicted and true value of the label. The choice of *loss function* depends on the nature of the problem at hand.

In regression tasks, where the goal is to predict a real-valued quantity, the two most commonly used *regression loss functions* are L_1 and L_2 loss.

L_1 loss function is also known as Mean Absolute Error (MAE) and it quantifies the average absolute error between true y_i and predicted $\hat{y}_i$ values,

$$L_{1_{loss(MAE)}} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|. \tag{3.4}$$

L_2 loss function, also known as Mean Squared Error (MSE), quantifies the average squared error between true y_i and predicted $\hat{y}_i$ values,

$$L_{2_{loss(MSE)}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2. \tag{3.5}$$

Due to its squaring operation, it is more sensitive to larger errors than smaller ones. In other words, the L_2 loss encourages the model to reduce the influence of outliers during training, hence minimizing their impact.

In multi-label classification, where a data point can be associated with multiple class labels, the Cross-Entropy loss (CE) is preferred. It measures the dissimilarity

between predicted class probabilities $\hat{y}_i$ and true class labels y_i. The CE loss for a single data point is defined as,

$$L_{CE} = -\sum_{i=1}^{C} y_i log(\hat{y}_i), \tag{3.6}$$

where C is the number of classes.

A natural approach to train multi-label classification models is by using the *Binary Cross-Entropy* (BCE) cost function. The BCE or *log loss* is also preferred for binary classification tasks, where a data point can be associated with one of the two possible classes. It is defined as

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} y_i ln(\hat{y}_i) + (1 - y_i) \cdot ln(1 - \hat{y}_i), \tag{3.7}$$

giving the average *binary cross-entropy* with y_i being the true binary label 0 or 1 and $\hat{y}_i$ the predicted probability with $\hat{y} \in [0, 1]$. When the true label is positive ($y = 1$) the loss encourages the model to assign high predicted probabilities ($\hat{y} \to 1$), and low predicted probabilities ($\hat{y} \to 0$) when the true label is 0. At the output layer of the neural network, the *sigmoid activation function* (see Section 3.1.1) is typically used to achieve a good performance.

Gradient descent

Once the loss function is established, the subsequent stage involves employing a training algorithm to decrease the loss. In simpler terms, the objective is to find a set of weights and biases that minimize the loss to the greatest extent possible. This can be achieved by using an algorithm known as *gradient descent* [23].

Suppose that L is a function of m variables $v_1, v_2, \dots, v_m$. Then the change ΔL is produced by a small change Δv as,

$$\Delta L \approx \nabla L \cdot \Delta v. \tag{3.8}$$

The gradient of loss L is defined as the vector of the partial derivatives,

$$\nabla L \equiv \left(\frac{\partial L}{\partial v_1}, \dots, \frac{\partial L}{\partial v_m}\right)^T \tag{3.9}$$

where T is the transpose operation. In order to guarantee the approximate expression 3.8 to be negative the choice of Δv to be written as

$$\Delta v = -\eta \nabla L \tag{3.10}$$

is made, where η is a small, positive parameter known as *learning rate*. Substituting this into the Equation 3.8 results to $\Delta L \approx -\eta ||\nabla L||^2$, which suggests $\Delta L < 0$. In this way L always decreases. Subsequent to the Equation 3.10 is the movement of the position v by the amount,

$$v \rightarrow v_m = v_{m-1} - \eta \nabla L_{m-1}. \tag{3.11}$$

By repeatedly applying this update rule L will continue to decrease until it reaches a global minimum. Essentially, the update rule (Eq. 3.10) defines the *gradient descent algorithm.*

In neural networks the *gradient descent algorithm* is applied to find the *weights* w and *biases* b that minimize the *loss* L. The aforementioned "position" v can be replaced with the parameters w_k, b_l. Hence, the gradient vector is now defined as

$$\nabla L(w_k, b_l) = \left(\frac{\partial L}{\partial w_k}, \frac{\partial L}{\partial b_l} \right)^T. \tag{3.12}$$

The change ΔL can be written as the approximation

$$\Delta L(w_k, b_l) \approx \frac{\partial L}{\partial w_k} \Delta w_k + \frac{\partial L}{\partial b_l} \Delta b_l. \tag{3.13}$$

Finally, the gradient descent update rule using the components w_k, b_l is

$$\begin{aligned} w_k \rightarrow w_k &= w_{k-1} - \eta \frac{\partial L}{\partial w_{k-1}} \\ b_l \rightarrow b_l &= w_{l-1} - \eta \frac{\partial L}{\partial b_{l-1}} \end{aligned}. \tag{3.14}$$

Hopefully, by repeatedly applying this rule the weights and biases keep updating until the minimum of the loss function is found. It is implied that the gradient ∇L needs to be computed separately for every input and then be averaged to give the true gradient. Unfortunately, this way is very time consuming and the learning is slow.

To speed up the learning process the *stochastic gradient descent* (SGD) can be used. Essentially, SGD uses a small number m of random training inputs x, called mini-batch. The resulting true gradient ∇L is the average value of the summation of all the gradients ∇L_{x_j} computed for the mini-batches. That is

$$\nabla L \approx \frac{1}{m} \sum_{j=1}^{m} \nabla L_{x_j}. \tag{3.15}$$

Applying this estimation to the *gradient descent* update rule (Eq. 3.14) for the *weights* w_k and *biases* b_l results to,

$$\begin{aligned} w_k \rightarrow w_k &= w_{k-1} - \frac{\eta}{m} \sum_j \frac{\partial L_{x_j}}{\partial w_{k-1}} \\ b_l \rightarrow b_l &= b_{l-1} - \frac{\eta}{m} \sum_j \frac{\partial L_{x_j}}{\partial b_{l-1}} \end{aligned}, \tag{3.16}$$

where the sums are over the training samples x that belong to the mini-batch. The *gradient descent* update rule is repeatedly applied until all mini-batches are

exhausted, hence completing an *epoch* (complete pass through the training data) of training. Then, this learning process starts over for a new *epoch*.

An extension of the SGD is the *Adam* optimization algorithm [24]. The name *Adam* derives from adaptive moment estimation. It is widely used in training neural networks due to its efficiency and low memory requirements. Adam is well-suited for problems that involve a large amount of data and/or a high number of parameters. The algorithm is straightforward to implement and its hyperparameters typically require little tuning.

To implement the *Adam* algorithm it is required to initialize the model parameters, *weights* and *biases*, and set the hyperparameters such as the *learning rate* η, the exponential decay rates for the moment estimates β_1, β_2 with $\beta_1, \beta_2 \in [0, 1)$, and a small constant ϵ to prevent division by zero. It is also required to initialize the two moment vectors m_0, v_0. The first moment (mean) m_0 keeps track of the exponentially decaying average of past gradients and the second raw moment (uncentered variance) v_0 keeps track of the exponentially decaying average of past squared gradients. They are both initialized as vectors of zeros. This can result in a slower learning process during the initial stages of optimization since it fails to capture important gradient information. This is particularly problematic when the learning rates or decay rates are set to small values. This issue can be counteracted by updating the bias moment estimates as,

$$\begin{aligned} m_t &\leftarrow \beta_1 m_{t-1} + (1 - \beta_1) \cdot g_t \\ v_t &\leftarrow \beta_2 v_{t-1} + (1 - \beta_2) \cdot g_t^2 \end{aligned}, \tag{3.17}$$

where g_t are the gradients of the loss function f evaluated with respect to the model parameters θ for each mini-batch t as $g_t = \nabla_\theta f_t(\theta)$. The next step is to compute the bias-corrected estimates by dividing m_t and v_t by factors of $(1 - \beta_1^t)$ and $(1 - \beta_2^t)$, respectively:

$$\begin{aligned} \hat{m}_t &\leftarrow \frac{m_t}{(1 - \beta_1^t)} \\ \hat{v}_t &\leftarrow \frac{v_t}{(1 - \beta_2^t)} \end{aligned}. \tag{3.18}$$

Finally, update the model parameters θ using the bias-corrected estimates resulting to the update rule of the *Adam* optimizer,

$$\theta_t \leftarrow \theta_{t-1} - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}. \tag{3.19}$$

Back-propagation algorithm

Back-propagation relies on the chain rule of calculus to compute gradients of the loss layer by layer [23]. It aims to compute the partial derivatives $\left(\frac{\partial L}{\partial w_k}, \frac{\partial L}{\partial b_l}\right)$ of the loss for all weights and biases in the network. However, before calculating these derivatives directly, an intermediate quantity δ_j^l needs to be introduced. Specifically, δ_j^l is the *error* in the j-th neuron in the l-th layer. Using the backpropagation algorithm δ_j^l is

computed for each layer of the network starting from the output layer and moving backward through the layers of the network. The *error* δ_j^l represents how much the predicted output deviates from the true value. Once δ_j^l is computed for each neuron, backpropagation establishes a relationship between δ_j^l and the partial derivatives $\frac{\partial L}{\partial w_k}, \frac{\partial L}{\partial b_l}$.

The *error* in the output layer L for every component j is given by

$$\delta_j^L = \frac{\partial L}{\partial \alpha_j^L} \sigma'(z_j^L), \tag{3.20}$$

where the term $\frac{\partial L}{\partial \alpha_j^L}$ measures the rate of change of the loss function in respect to the activation α_j^L of the j-th neuron in the output layer. The second term on the right $\sigma'(z_j^L)$ measures how fast the activation function σ is changing concerning the input z_j^L. The value z_j^L is defined as the weighted input to layer l: $z_j^L \equiv w^l \alpha^{l-1} + b$, where α^{l-1} represent the activation on the previous layer and w^l, b are the respected *weights* and *biases*, and is already computed during the forward pass.

Rewriting the Equation 3.20 in a matrix-based form using the *Hadamard product* $\odot$, an element-wise vector multiplication,

$$\delta^L = \nabla_\alpha L \odot \sigma'(z^L), \tag{3.21}$$

where $\nabla_\alpha L$ is defined as a vector with components the partial derivatives $\frac{\partial L}{\partial \alpha_j^L}$. The *error* δ^l in a layer l in terms of the *error* δ^{l+1} of the preceding layers $l+1$ is given by

$$\delta^l = ((w^{l+1})^T \delta^{l+1} \odot \sigma'(z^l). \tag{3.22}$$

The transpose of the weight matrix (w^{l+1}) of the $(l+1)$-th layer is used to propagate the error backward through the network, moving from the output layer $l+1$ back to the previous layer l. It is applied to the *error* δ^{l+1}, essentially moving the *error* backward giving a measure of how much the error at the $(l+1)$-th layer contributes to the error at the output of the l-th layer. The Hadamard product $\odot\sigma'(z^l)$ moves the error backward through the activation function in layer l. Thus, calculating the error δ^l at the weighted input to each layer l. Finally, the changes of loss (gradient) with respect to any bias b_j^l and any weights w_{jk}^l in the network can be calculated with the equations

$$\frac{\partial L}{\partial b_j^l} = \delta_j^l, \quad \frac{\partial L}{\partial w_{jk}^l} = \alpha_k^{l-1} \delta_j^l. \tag{3.23}$$

Common approaches to improve the neural network

There are several ways to improve the performance of a neural network. This section highlights some of them.

Firstly, the hyperparameter *learning rate* needs to be tuned depending on the algorithm. As mentioned in Section 3.1.2 the learning rate is applied to the gradient and scales it. It can be considered to be similar to a step size. Its value needs to be small enough to give a good approximation yet not too small, which could result in

slow convergence or confinement to a local minima. Conversely, if the value is too large the algorithm might fail to converge as it could overshoot minima. A viable approach is to decrease the learning rate as the number of *epochs* progresses.

Secondly, regularization techniques can be applied to prevent overfitting [25]. Regularization refers to adjustments made to a learning algorithm with the aim of decreasing its generalization error while keeping its training error unchanged. *Overfitting* occurs when a model excels at predicting labels for the training examples but makes errors when applied to unseen data that were not part of the training dataset. The most common techniques are *dropout* ,*early stopping* and *batch-normalization*. *Dropout* is a straightforward concept in neural networks. During the training process, for each training example, random neurons in the network are temporarily excluded from the computation. The extent of dropout is controlled by a parameter, usually in the range $[0, 1]$, where a higher value implies a more significant regularization effect. Dropout can be implemented in the neural network by inserting a dropout layer between two consecutive layers, or by specifying the dropout parameter for a specific layer. *Early stopping* involves periodically evaluating the performance of the model on a validation dataset during training. After each epoch, the current state of the neural network model is saved as a checkpoint. The model performance is assessed on a separate validation dataset. If the model performance on the validation set begins to deteriorate it is an indication that the model is starting to overfit the training data. Once a decreased performance on the validation set is observed the training can be stopped. *Batch normalization* is a method used to normalize the weighted outputs of each layer before passing them as input to the next layer. It can be incorporated in the network architecture by inserting a batch normalization layer between two consecutive layers. Technically, it is not a regularization technique but it has regularization effect on the model.

Thirdly, a common technique used to enhance the network performance is *pooling*. In this technique, a pooling layer takes each feature map output from the previous layer and produces a more compact feature map, as shown in Figure 3.3. Pooling is essentially a downsampling process that reduces the spatial dimensions by half. For instance, in average pooling, the features' average values are calculated and used as the output. This approach offers the significant advantage of yielding fewer pooled features, thereby reducing the number of parameters required in subsequent layers.

Finally, the problem of *vanishing gradients* needs to be addressed. As mentioned in Section 3.1.2, during the process of training a neural network using gradient descents, the network parameters, which include weights w and biases b, are adjusted in small steps to minimize the loss function. This adjustment is determined by calculating the gradient of the loss function with respect to each parameter. However, in some situations, this gradient can become extremely small, approaching zero. When gradients are too small, the parameters associated with those gradients stop updating, thus no learning is achieved. To prevent this from the start of the training process, the weights are sampled from a Gaussian distribution, where the width is inversely proportional to the number of connections of the neuron

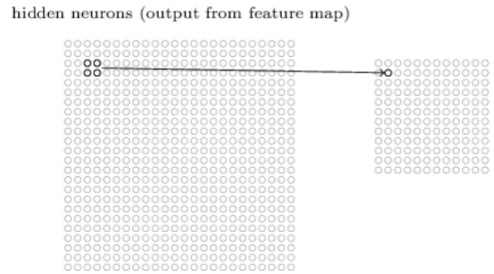


Figure 3.3: This diagram demonstrates an instance of how the pooling technique operates. On the left side, it displays the output of a layer containing 24x24 neurons, while on the right, 12x12 neurons remain after the pooling process. Modified from [23].

and centered around zero. Additionally, in the subsequent training phases, the batch-normalization technique, as mentioned earlier, is employed.

3.2. Graph Neural Networks

In particle physics, there is a growing interest in leveraging deep learning as an exceptional resource to tackle intricate scientific challenges that extend beyond basic classification. It serves as a means to effectively structure and comprehend extensive datasets, derive conclusions about unobserved causal factors, and even discover fundamental physical principles governing complex phenomena. Notably, within High Energy Physics (HEP), machine learning is frequently employed to decipher complex inverse relationships, with the goal of deducing insights into the underlying physics processes based on information acquired from detectors.

Graph Neural Networks (GNNs) are a specific class of neural network architecture designed to operate on graph-structured data. By definition, data are sets of items. When examining relationships between these items, whether geometric or physical, a set can be transformed into a graph by incorporating an adjacency matrix. Physics data is collected from experiments, where each data point corresponds to a specific event or measurement. They can be thought of as "points" in space, each having specific attributes associated with them. Point clouds are a collection of data points represented in 3D space in computer vision and often used in deep-learning algorithms.

The data collected from the KM3NeT detector have many features of point clouds, making it a logical choice to exploit graph neural networks for the analysis.

Some advantages on the usage of graphs with respect to image based methods (e.g CNN) are linked to the limited resolution on position and time that can be achieved through images/fixed grid pixels. In principle bin size can be increased to reach any desired resolution, but this operation would cause the increase of the image size, of memory usage and sparsity of the signal. At the same time, DOMs in KM3NeT detector are moving under the effect of sea current. This information is therefore lost inside the position bin size. These techniques have been also

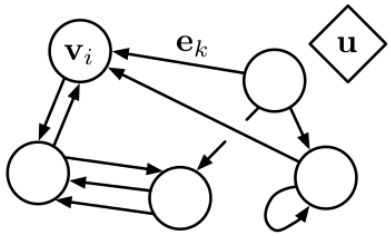


Figure 3.4: Representation of a directed graph consisting of a global feature u, nodes v_i, and edges e_k connecting the nodes pairwise. Figure from [26].

developed in the context of KM3NeT experiment. However, the fixed structure of pixel has shown limitations in the capability to represent data collected by the telescope. The high-dimensional and sparse signal registered in the detector can be much better encoded in graphs.

3.3. Graph Neural Networks in KM3NeT

3.3.1. Structure of graph neural networks

The input data are expressed as graphs, conforming to the architecture [26] of

- a global feature $\boldsymbol{u}$, representing the entire graph and carrying some physical meaning (particles etc),
- the nodes $\boldsymbol{v}_i$, which are the data points possessing multiple attributes (position, time etc) and
- the edges $\boldsymbol{e}_k$, connecting pairs of nodes and defining the relationship between them (distance etc).

An illustration of this can be seen in Figure 3.4. The edges are connecting the nodes one way, allowing the flow of information between the nodes in certain directions. For memory usage, optimization and for keeping the number of connections under control, each node is connected to its *k-nearest neighbours*. This is typically selected by the geometrical distance between the nodes. The internal formalism of how a graph $\boldsymbol{u}$ is updated is shown in Figure 3.5. The stages of processing are the following,

$$\begin{aligned} \boldsymbol{e}'_{\boldsymbol{k}} &= \phi^e(\boldsymbol{e}_{\boldsymbol{k}}, \boldsymbol{v}_{r_k}, \boldsymbol{v}_{s_k}, \boldsymbol{u}) & \bar{\boldsymbol{e}}'_{\boldsymbol{i}} &= \rho^{e\to v}(E'_i) \\ \boldsymbol{v}'_{\boldsymbol{i}} &= \phi^v(\bar{\boldsymbol{e}}'_{\boldsymbol{i}}, \boldsymbol{v}_{\boldsymbol{i}}, \boldsymbol{u}) & \bar{\boldsymbol{e}}' &= \rho^{e\to u}(E') \\ \boldsymbol{u}' &= \phi^u(\bar{\boldsymbol{e}}', \bar{\boldsymbol{v}}', \boldsymbol{u}) & \bar{\boldsymbol{v}}' &= \rho^{v\to u}(V') \end{aligned} \tag{3.24}$$

where ϕ^e, ϕ^v, ϕ^u are the three *update functions* and $\rho^{e\to v}, \rho^{e\to u}, \rho^{v\to u}$ are the three *aggregation functions*. Initially, for each edge $\boldsymbol{e}'_{\boldsymbol{k}}$ an output edge is computed and all

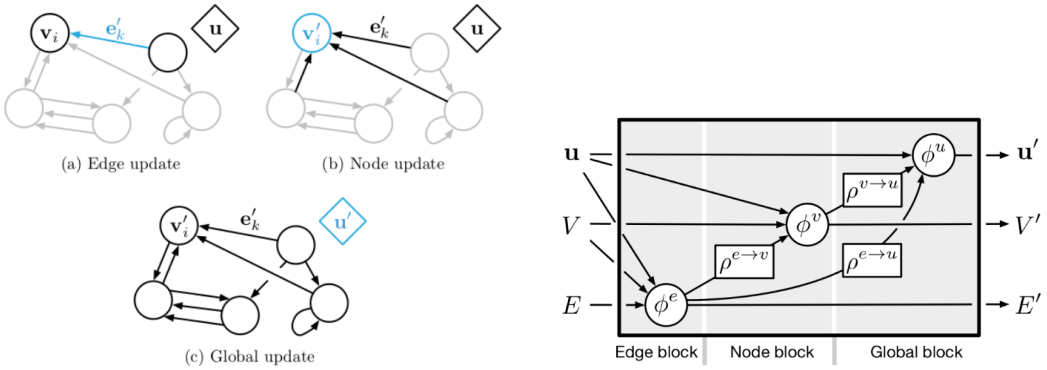


Figure 3.5: Left: Illustration of how a graph is updated. Starting with updating the edges (a), which are then aggregated to a node (b), concluding to the update of the entire graph (c). Figure taken from [26]. Right: Representation of a GNN block. The input graph $G = (u, V, E)$ is processed using the update/aggregation functions and the output graph $G' = (u', V', E')$ with updated attributes is returned. Figure taken from [27].

the output edges are aggregated to a node $\boldsymbol{e}'_i$, where E'_i represents the set of edges in the i-th node. Then, the nodes are updated yielding the output nodes $\boldsymbol{v}'_i$. The updated edges and nodes are all aggregated resulting to the new graph with global attribute $\boldsymbol{u}'$. This formalism is a general framework that can be used in several GNN architectures.

Loss function

In this study, the loss function used for the classification tasks is the *Binary Cross Entropy*, as defined in Section 3.1.2. For the regression tasks, a log-normal distribution is preferred. The loss is given by,

$$L = ln\sqrt{2\pi} + ln(\sigma) + \frac{(y-\mu)^2}{2\sigma^2}, \tag{3.25}$$

whereas for the minimization only the terms that contain the width σ and the mean value μ are of interest

$$L = ln(\sigma^2) + \frac{(y-\mu)^2}{\sigma^2}. \tag{3.26}$$

For the special case of $\sigma = 1$ the Eq. 3.26 results to the mean squared error as defined in Eq. 3.5.

3.3.2. Edge convolution

The edge convolution (EdgeConv) operation [28] has to be defined, which is an approach of learning on point cloud data. In EdgeConv the point cloud is represented as a graph, where the vertices are the points (nodes) and the edges are the connections between these points to its k-nearest neighbors. The edge features are defined as $\boldsymbol{e}_{ij} = \boldsymbol{h}_{\Theta}(\boldsymbol{x}_i, \boldsymbol{x}_j)$, where $\boldsymbol{h}_{\Theta} : \mathbb{R}^F x \mathbb{R}^F \rightarrow \mathbb{R}^{F'}$ is a function with learnable

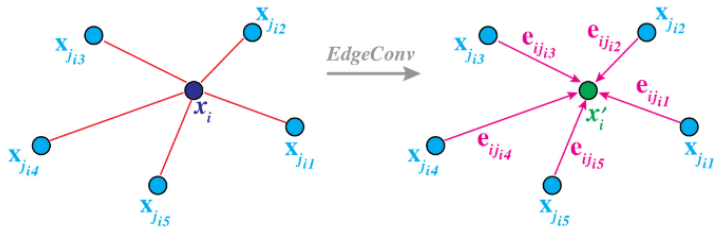


Figure 3.6: The EdgeConv operation. The central point x_i is surrounded by a local patch consisting of points x_{i_j}, which illustrates the connections of a node with its k nearest neighbors. The right side shows the output x_i' of the EdgeConv, calculated by aggregating the edge features linked to all the edges originating from each connected point. Figure from [28].

parameters Θ and the dimensions F represent the feature dimensionality of a given layer. The output of the EdgeConv operation for each point x_i has the form

$$x_i' = \square_{j=1}^{k} h_\Theta(x_i, x_j), \tag{3.27}$$

where $\square$ is a symmetric aggregation operation (sum, max etc). The best option for the operation $\square$ is the mean, i.e $\frac{1}{k}\sum$. For the purposes of this analysis the asymmetrical edge function is preferred, that is

$$h_\Theta(x_i, x_j) = \bar{h}_\Theta(x_i, x_j - x_i) \tag{3.28}$$

Overall, when EdgeConv is applied to a point cloud with F dimensions and n points, it produces a point cloud with F' dimensions, retaining the same number of points. What's particularly noteworthy is its ability to be easily stacked. This stacking takes one point cloud and converts it into another with the same number of points, only changing the dimensions of the feature vector for each point. This stacking capability allows the creation of deep networks that can progressively learn features from point clouds in a hierarchical manner. Furthermore, the feature vectors derived from EdgeConv can be interpreted as new coordinates within a latent space. Consequently, distances between points to be calculated in this latent space, facilitating the adaptive learning of point proximity through EdgeConv operations.

3.3.3. ParticleNet architecture

For what concerns the model architecture adopted for the analysis proposed in the next sections, the ParticleNet architecture has been exploited, as defined in [1]. The overall structure can be seen in the left side of Figure 3.7. The network consists of three stacked Edge Convolutional layer blocks, a global average pooling layer and two fully connected layers. The input is comprised of two quantities, coordinates and features. Coordinates are only involved in determining k-nearest neighbors (k-NN) and comprise hit information related to $ct, x-, y-, z-$ position, measured using Euclidean distance. Features encompass information for each node and edge. Node features consist of all pertinent hit information (time, 3d position,

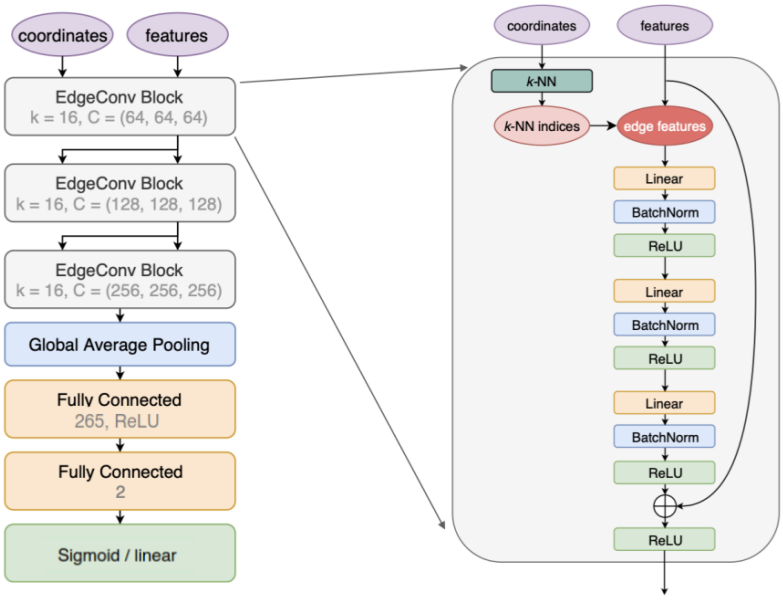


Figure 3.7: Left: Diagram of the ParticleNet architecture. Right: Close up of the inner structure of an Edge Convolutional block. Modified from [1].

3d direction), including the direction of the PMT. Edge features, denoted as e_{ij}^k, represent the difference between adjacent nodes v_i and v_j, i.e $e_{ij}^k = v_i - v_j$, where $k \in$ {time, x-, y-, z-position, x-, y-, z-direction}.

The right side of Figure 3.7 provides a more detailed view of the edge convolution block. In the first EdgeConv block the k-nearest neighbors are calculated for updating the node. This block uses the the spatial coordinates of the particles to compute the distances. A small kernel network, comprised of three layers with 64 neurons each, is trained based on the input features. This kernel network linearly processes the input, applies batch normalization, and employs the ReLU activation function. The contributions from all nodes and edges to this kernel network are averaged for a single training sample, forming the convolution operation. At the end of the block, an updated graph is produced, featuring newly defined nodes and edges. In the output graph, the original features are also retained and can be partially combined with the updates.

The subsequent two EdgeConv blocks follow the same layout, however the number of neurons per layer increase to 128 and 256 to accommodate the storage of higher-level features. Each EdgeConv block receives as input the graph in latent space from the preceding block and apply the same operations. This approach ensures that the graph evolves as it passes through the network, allowing for the repeated redefinition of edges. After the last EdgeConv block, the average pooling layer is applied to reduce the dimensions. Then, the 256 features are supplied to the fully connected layer, which uses the ReLu activation function. A second fully connected layer with two nodes, followed by a function, depending on the specific task, is used to generate the output.

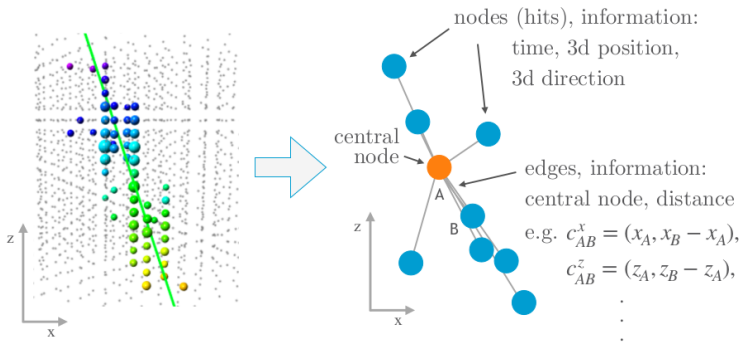


Figure 3.8: Presentation of the KM3NeT data to graphs. Left: Illustration of a muon track (green line) in the KM3NeT detector. Right: The resulting graph, where the nodes hold the hit information (3d position, 3d direction, time). In the central node (orange) the information from all the edges is aggregated. Figure taken from [7].

This architecture was originally designed for point cloud applications, and used for jet tagging at LHC, showing outstanding performances with respect to image convolutional techniques.

The KM3NeT detector can be modelled as a 3D array of photo-sensors, called Digital Optical Module (DOM), capable to register the arrival time and time over threshold of the photon hits impinging on one of the 31 PMTs contained inside a DOM. Every event registered in the KM3NeT telescope can therefore be seen as a set of photon hits, for which time and coordinates are known. The most natural way to encode information of events into a graph is to see every photon hit as a node. In turn, each node has a 7-dimensional feature space represented by: 3 spatial coordinates, 3 directions of the PMTs and time. Each connection between two nodes, will be characterized by a $2 \times F$ dimensional feature space consisting of: the feature of the starting node itself and the difference with respect to the neighbour node to which the starting node is connected to (see Fig. 3.8). To create the final graph structures then, nodes should be connected to each other, defining an arbitrary distance metric. Since the features of each node are positions and times of the photon hits, a natural choice for the distance is the Euclidean. At this point the convolution operation (see Sec. 3.6) is done applying a multilayer Percepton network (called a kernel network from now on) on the $2\times$ F dimensional feature space of the edges, calculated previously, and producing as result an updated vector of features. The kernel network is slid over the edges of the graphs forming the event. The final output of the convolution layer will be then the average length of the updated vectors of k nearest neighbours of each specific node. The kernel network used in this work is a multi-layer perceptron (MLP) consisting of three dense layers and three corresponding batch normalization layers.

3.3.4. Software

The analysis software employed in this study is a collaborative effort between the KM3NeT Collaboration and the broader deep learning community.

The initial data processing is accomplished using a combination of orcasong [29] and km3pipe [30]. This pre-processing stage entails converting the input ROOT files into h5 files using *km3pipe* and subsequently generating the graphs with *orcasong*. The neural network training and evaluation are conducted using orcanet [31], a framework built upon the Keras and Tensorflow libraries.

4

Event classification and energy reconstruction with graph neural networks

The performance of the Graph Neural Network is studied for particle identification and energy regression. Multiple networks have been trained and evaluated using KM3NeT/ARCA data for the three following tasks:

- Signal/background classification. The network is trained to distinguish between signal (neutrinos) and background (atmospheric muons), yielding a neutrino probability.
- Track/shower classification. The network learns to classify the events into tracks or showers, yielding a track probability.
- Energy regression. The network predicts the energy of the interaction, yielding the predicted energy and its uncertainty.

All networks are trained on a single GPU. As mentioned before, the raw data received from either the real detector or a simulated one need to be preprocessed to be used from the GNN. The hit information of an event are converted into graph structures. In these graphs, the nodes are the PMT hit coordinates in space, direction, ToT and time and the edges connect the neighboring nodes. To train a neural network on such data, it is crucial to provide a proper training set for each application. For classification tasks, it is essential to maintain a balanced dataset between the different categories. For regression, the value of the target variable is important.

For the two classifiers the true labels that serve as the basis for the networks to learn from and compute the loss are directly determined from the particle, i.e a $\nu_\mu CC$ is assigned 1 in the signal/background classifier and 1 in the track/shower classifier, whereas a ν_eCC is assigned 1 in the signal/background classifier and 0 in the track/shower classifier. The loss is calculated form the *Binary Cross Entropy* as defined in Section 3.1.2. In the case of the energy reconstruction, the decadic logarithm of the energy, $log_{10}E$, is the true label and the loss is derived from the difference between predicted and true value as well as the predicted uncertainty, as mentioned in Section 3.3.1.

The terms *shower* and *track* neutrinos, as mentioned in Section 1.6.1, are used to characterize the neutrino interactions that produce different signatures on the detector. Track neutrinos produce a muon in their outgoing channel, such as muon (anti) neutrino charged-current (CC) interactions (i.e $\nu_\mu CC, \overline{\nu}_\mu CC$). Shower neutrinos are electron (anti) neutrino CC interactions and all neutral-current (NC) interactions (i.e $\nu_e CC, \overline{\nu}_e CC, \nu_e NC, \overline{\nu}_e NC, \nu_\mu NC, \overline{\nu}_\mu NC$). In general, shower events emit their light within a relatively small volume, which implies that they are either predominantly confined within the detector or, when located at a greater distance from the detector, may not trigger at all.

To assess the progress of the training, the curves for the loss and accuracy are plotted over the course of training epochs for both validation and training sets. The final model used for the analysis is the one with the lowest validation loss for each task. For the model architecture adopted in the next sections, the ParticleNet architecture has been exploited [1].

Selection criteria

To suppress events generated by environmental optical background, mainly due to ^{40}K decay, the following selection criteria are applied to each event.

- The minimum number of triggered hits is requested to be 25, thus rejecting ambiguous events with only a few signal hits and reducing the atmospheric muon contamination.
- Selecting only upgoing events by accepting events reaching the detector from zenith angles greater than 90 deg to further suppress the atmospheric muons.
- The minimum value of log-likelihood is requested to be greater than 50 to discard events with very low probability to be signal.
- Lastly, a minimum uncertainty of the energy reconstruction is set to a factor of 1 to cast outliers.

4.1. Signal/Background classification

A classification model has been trained to distinguish between atmospheric muons and neutrinos. The classifier produces a score for each event, ranging from 0 to 1, that represents the probability the network assigns to that event to be of a certain class. During the training phase the Graph Neural Network (GNN) takes as input the graphs created from the Monte Carlo (MC) simulation. The training process utilizes approximately 90% of the dataset for training and 10% for validation.

Specifically, three training sessions were conducted, with one utilizing ARCA7 v8.0 MC and the other two employing ARCA8 v8.0 MC data as input with an energy spectrum that ranges from 10^2 to 10^8 GeV for the neutrinos. The datasets consist of 500K, 1 million events, and 2 million events respectively, equally divided among atmospheric muons and neutrinos. All networks are trained with ReLU

as activation function, Adam as optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 0.1$), epochs equal to 20, initial learning rate 0.03 with a 0.1% decrease after each epoch, batchsize 32 and k-nearest neighbors (k-NN) equal to 10 (ARCA8) and 16 (ARCA7, ARCA8). The evaluation of the network trained on ARCA7 involved analyzing a period of 25.5 days, while for the network trained on ARCA8, a period of 170 days was examined. The analysis results for each training session are presented in the following sections. Additionally, the evaluation of the ARCA6 trained network was performed making predictions on ARCA7 v8.0 MC and ARCA8 v8.0 MC data samples. The analyzed period consists of 25.5 and 22.2 days, respectively. The training session using ARCA6 v6.3 MC data as input was conducted by Alba Domi, using the same classification model and hyperparameters.

4.1.1. GNN performance on ARCA6

In this section, the performance of the ARCA6 trained network is examined on the ARCA7 and ARCA8 data. In the following plots, the distributions of neutrino scores with weighted events (to account for the atmospheric and cosmic neutrino flux) are presented on the upper part and the comparison between the inference on data over inference on MC simulation on the lower part. In Figures 4.1, 4.2 the results for the ARCA6 trained network evaluated on the ARCA7 data are presented. In Figure 4.1 no selections are applied to the events, whereas the plots in Figure 4.2 are generated by applying the selections defined on 4, i.e on the number of triggered hits for each event $n\_trig\_hits > 25$, on the log-likelihood $log\_lik > 50$ and by selecting only the upgoing events with $zenith > 90$. These selections are applied to suppress events generated by environmental optical background, mainly due to ^{40}K decay. Similarly, using the ARCA8 data for the inference, Figure 4.3 shows the probability distribution and the data-MC comparison without selections, whereas in Figure 4.4 the predefined selections are applied. For both inferences, after implementing the selections, a good agreement between data and MC is evident up to nu_score = 0.8.

To further assess the performance of the GNN-trained networks, the results for each trained network inferred with each dataset are displayed in the tables below, showcasing the application of a selection for low and high neutrino scores. The initial three rows exhibit the results without any additional selections, while the subsequent rows demonstrate the impact of applying a selection based on triggered hits, log-likelihood and zenith. The evaluation results of the ARCA6 trained network, using ARCA7 and ARCA8 data, are presented in Table 4.1. Focusing on high neutrino scores (nu_score> 0.98), only 1.82% and 0.95% of atmospheric neutrinos remain in the respective data-sets, as well as 14.28% and 11.43% of cosmic neutrinos. This indicates that the classifier struggles to effectively differentiate between neutrinos and the background. Atmospheric muons are correctly assigned a neutrino score closer to 0, with only 0.08% and 0.01% of muons remaining at nu_score> 0.98.

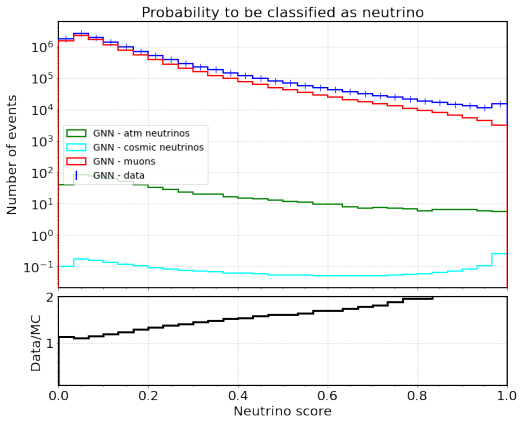


Figure 4.1: Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio. **ARCA6** trained network using ARCA7 data for evaluation.

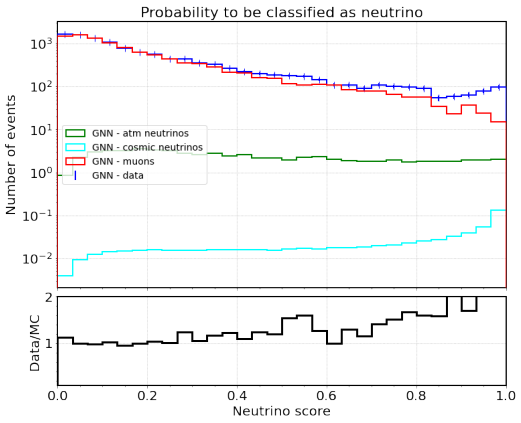


Figure 4.2: Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio. Selection criteria: $n\_trig\_hits > 25$, $log\_lik > 50$, $zenith > 90$ (up-going). **ARCA6** trained network using ARCA7 data for evaluation.

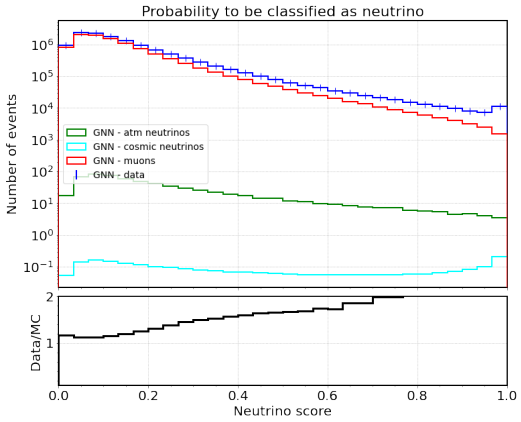


Figure 4.3: Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio. **ARCA6** trained network using ARCA8 data for evaluation.

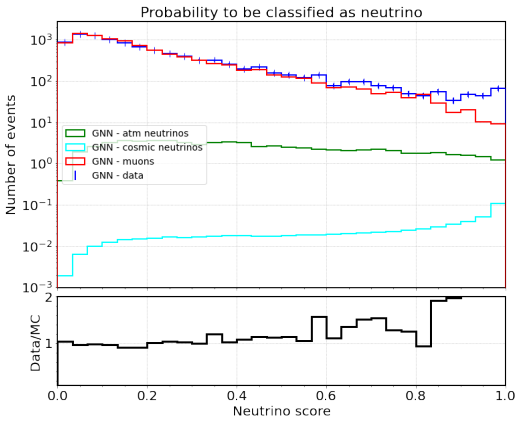


Figure 4.4: Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio. Selection criteria: $n\_trig\_hits > 25$, $log\_lik > 50$, $zenith > 90$ (up-going). **ARCA6** trained network using ARCA8 data for evaluation.

ARCA6 TRAINING - ARCA7 INFERRED			
	events w\o cuts	nu score<=0.2	nu score>=0.98
data	11759482 (100%)	9272786 (78.85%)	10736 (0.001%)
atm. muons	9624273.2 (100%)	7955951.5 (82.67%)	1705.8 (0.02%)
atm. neutrino	655.2 (100%)	351.7 (53.68%)	3.37 (0.51%)
cosmic neutrino	2.56 (100%)	0.80 (31.25%)	0.19 (7.42%)
trig. hits>25 & log. lik.>50 & zenith>90			
data	12344 (100%)	7564 (61.28%)	74 (0.60%)
atm. muons	11273 (100%)	7411.2 (65.74%)	9.08 (**0.08%**)
atm. neutrino	73 (100%)	17.60 (24.11%)	1.33 (**1.82%**)
cosmic neutrino	0.70 (100%)	0.07 (10.00%)	0.10 (**14.28%**)

ARCA6 TRAINING - ARCA8 INFERRED			
	events w\o cuts	nu score<=0.2	nu score>=0.98
data	12348906 (100%)	9501884 (76.94%)	8234 (0.0001%)
atm. muons	10090861 (100%)	8170448 (80.97%)	775 (0.008%)
atm. neutrino	674.35 (100%)	350.5 (51.98%)	1.96 (0.29%)
cosmic neutrino	2.54 (100%)	0.75 (29.53%)	0.16 (6.30%)
trig. hits>25 & log. lik.>50 & zenith>90			
data	10538 (100%)	6350 (60.26%)	46 (0.44%)
atm. muons	10288.7 (100%)	6596.2 (64.11%)	1.18 (**0.01%**)
atm. neutrino	76.67 (100%)	16.58 (21.62%)	0.73 (**0.95%**)
cosmic neutrino	0.70 (100%)	0.06 (8.57%)	0.08 (**11.43%**)

Table 4.1: Left: Evaluation of the ARCA6 trained network using ARCA7 data. Right: Evaluation of the ARCA6 trained network using ARCA8 data.

4.1.2. GNN performance on ARCA7

	signal-background classifier
track neutrinos	226 (48.9%) / 47 (45.6%)
shower neutrinos	0
atm. muons	236 (51.1%) / 56 (54.3%)
total	462 / 103

Table 4.2: Number of events used in training and validation datasets for the signal/background classification network using ARCA7 data. The first value represents the absolute number of events in thousands in the training set and the second value the number in the validation set, with the proportion indicated in parentheses.

In this training session, ARCA7 v8.0 MC data are used as input consisting of approximately 560.000 atmospheric muons and track neutrino events. The training set consists of 460K events and the validation set of 100K events. An overview of the two sets is presented in Table 4.2. In Figure 4.5 the training and validation loss and accuracy curves are visualised. The loss decreases steadily as the training progresses. The decreasing validation loss indicates that the model generalizes well to unseen data. The model that is used for evaluating the trained network is the one with epoch equal to 20. The evaluation involved analysing a period of 25.5 days. The neutrino probability is presented in Figure 4.6, 4.7 where the distributions of neutrino scores with weighted events at the upper part and the ratios between the inference on data over inference on MC simulation at the lower part are shown. In Figure 4.6 no selections are applied to the events, whereas the plots in Figure 4.7 are generated by applying a selection on the number of triggered hits for each event

$n\_trig\_hits > 25$, on the log-likelihood $log\_likelihood > 50$ and by selecting only the up-going events with $zenith > 90$.

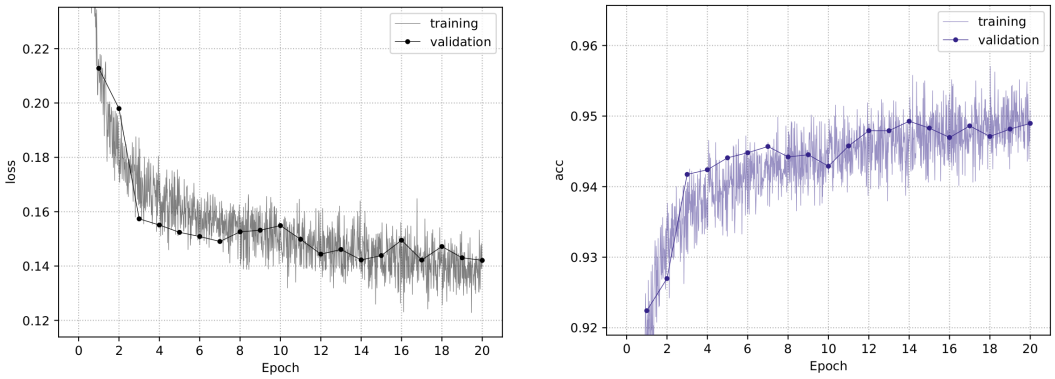


Figure 4.5: ARCA7, k-NN=16 training. Left: The loss curves for the training and validation (with dots) sample for ARCA7. Right: The accuracy curves for the training and validation (with dots) sample for ARCA7.

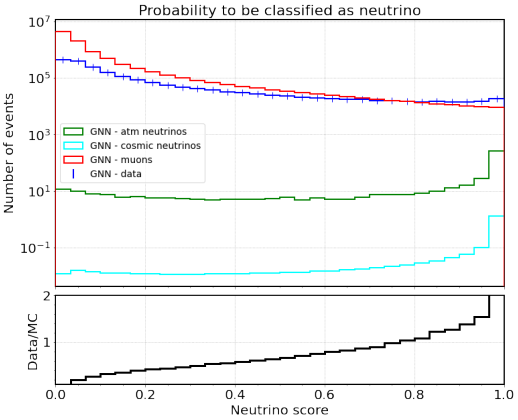


Figure 4.6: ARCA7 trained network using ARCA7 data for evaluation. Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio.

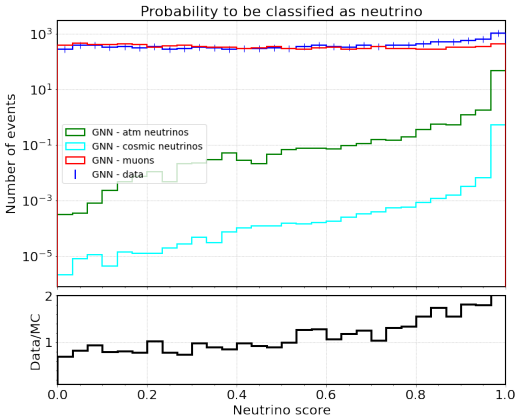


Figure 4.7: ARCA7 trained network using ARCA7 data for evaluation. Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio. Selection Criteria: $n\_trig\_hits > 25$, $log\_lik > 50$, $zenith > 90$ (up-going).

The evaluation results of the ARCA7 trained network, using ARCA7 and ARCA8 data, are presented in Table 4.3. Focusing on high neutrino scores (nu_score> 0.98), 85.8% and 78.47% of atmospheric neutrinos remain in the respective datasets, as well as 96.2% and 91.43% of cosmic neutrinos. Additionally, only 2.58% and 1.00% of atmospheric muons survive this selection. Overall, the performance of the ARCA7 network seems to result in a better classification when using data from 7 DUs.

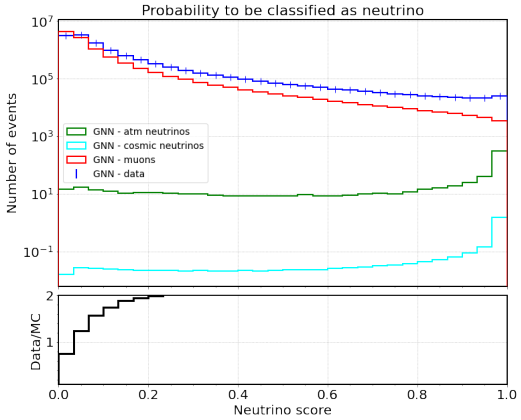


Figure 4.8: ARCA7 trained network using ARCA8 data for evaluation. Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio.

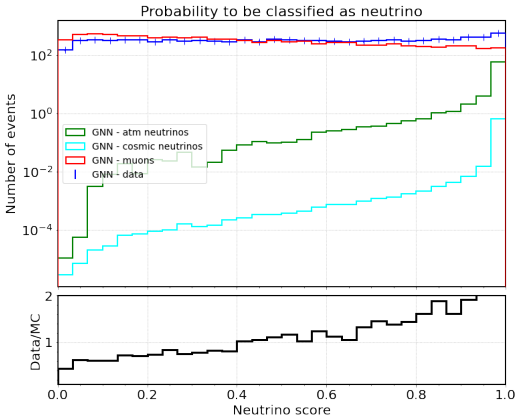


Figure 4.9: ARCA7 trained network using ARCA8 data for evaluation. Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio. Selection Criteria: $n\_trig\_hits > 25$, $log\_lik > 50$, $zenith > 90$ (up-going).

ARCA7 TRAINING - ARCA7 INFERRED			
	events w\o cuts	nu score<=0.2	nu score>=0.98
data	11759482 (100%)	9603854 (81.67%)	31256 (0.003%)
atm. muons	9221943.6 (100%)	8222680 (89.16%)	5299.5 (0.06%)
atm. neutrino	496.9 (100%)	49.1 (9.88%)	249.8 (50.3%)
cosmic neutrino	1.95 (100%)	0.08 (4.10%)	1.26 (64.62%)
trig. hits>25 & log. lik.>50 & zenith>90			
data	11668 (100%)	2020 (17.31%)	718 (6.15%)
atm. muons	10852 (100%)	2661.3 (24.52%)	280.5 (**2.58%**)
atm. neutrino	55.8 (100%)	0.02 (0.03%)	47.9 (**85.8%**)
cosmic neutrino	0.53 (100%)	0.0001 (0.01%)	0.51 (**96.2%**)

ARCA7 TRAINING - ARCA8 INFERRED			
	events w\o cuts	nu score<=0.2	nu score>=0.98
data	12348906 (100%)	10408152 (84.28%)	17372 (0.001%)
atm. muons	10090861 (100%)	9242488 (91.59%)	2014.57 (0.02%)
atm. neutrino	674.35 (100%)	79.6 (11.80%)	275.32 (40.83%)
cosmic neutrino	2.54 (100%)	0.14 (5.51%)	1.43 (56.30%)
trig. hits>25 & log. lik.>50 & zenith>90			
data	10538 (100%)	1946 (18.47%)	436 (4.14%)
atm. muons	10288.7 (100%)	3119.57 (30.32%)	102.51 (**1.00%**)
atm. neutrino	76.67 (100%)	0.04 (0.05%)	60.16 (**78.47%**)
cosmic neutrino	0.70 (100%)	0.0002 (0.03%)	0.64 (**91.43%**)

Table 4.3: Left: Evaluation of the ARCA7 trained network using ARCA7 data. Right: Evaluation of the ARCA7 trained network using ARCA8 data.

4.1.3. GNN performance on ARCA8

	signal-background classifier
track neutrinos	287 (43.6%) / 53 (27.5%)
shower neutrinos	0
atm. muons	372 (56.4%) / 140 (69.6%)
total	659 / 193

	signal/background classifier
track neutrinos	405 (23.4%) / 42 (22.3%)
shower neutrinos	392 (22.7%) / 38 (20.2%)
atm. muons	930 (53.8%) / 108 (57.4%)
total	1727 / 188

Table 4.4: Number of events used in training and validation datasets for the signal/background classification network utilizing ARCA8 data (k-NN=10) (left) and ARCA8 data (k-NN=16) (right). The first value represents the absolute number of events in thousands in the training set and the second value the number in the validation set, with the proportion indicated in parentheses.

Two networks are trained using ARCA8 v8.0 MC data as input. The effect of the hyperparameter k-nearest neighbors (k-NN) on GNN performance is examined. One network is trained with 1M events and k-NN equal to 10, and the other one with 2M events and k-NN equal to 16. The default setting for k-NN is 16, although an investigation was conducted with k-NN = 10. However, as demonstrated below, the results were not as favorable, a conclusion also supported by Lukas Hennig [32]. An overview of the two sets for both trained networks is presented in tables 4.4. The loss and accuracy plots for both networks are visualised in Figure 4.10 and Figure 4.11 respectively. At the network with k-NN=10, the expected steady decrease of the validation loss as the training progresses is not achieved. At the network with k-NN=16 the validation loss initially decreases but after epoch=12 it remains constant which is a sign of overfitting.

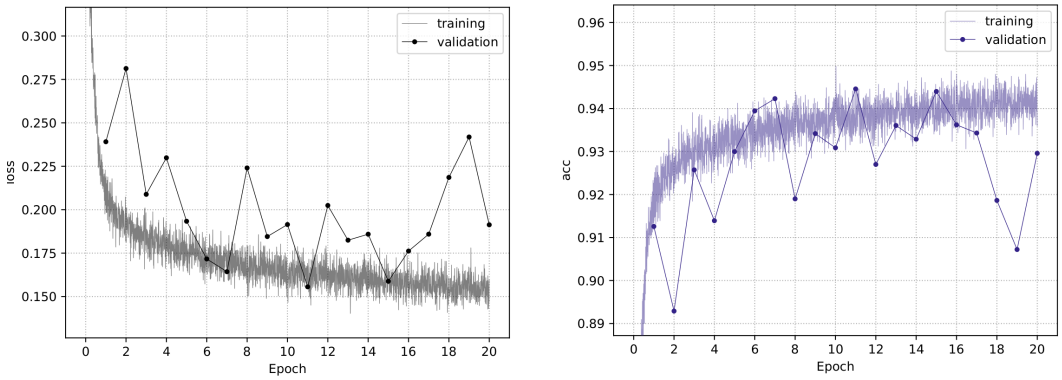


Figure 4.10: ARCA8, k-NN=10 training. Left: The loss curves for the training and validation (with dots) sample for ARCA8. Right: The accuracy curves for the training and validation (with dots) sample for ARCA8.

The models that are used to evaluate the networks are the ones with epoch equal to 11 and 19, respectively. The evaluation for both networks involved analyzing a period of 170 days. The analysis results are shown in Figures 4.12, 4.13 for the trained network with k-NN=10 and in Figures 4.14, 4.15 for the one with k-NN=16, where the probability for each event to be classified as neutrino and the data-MC comparison

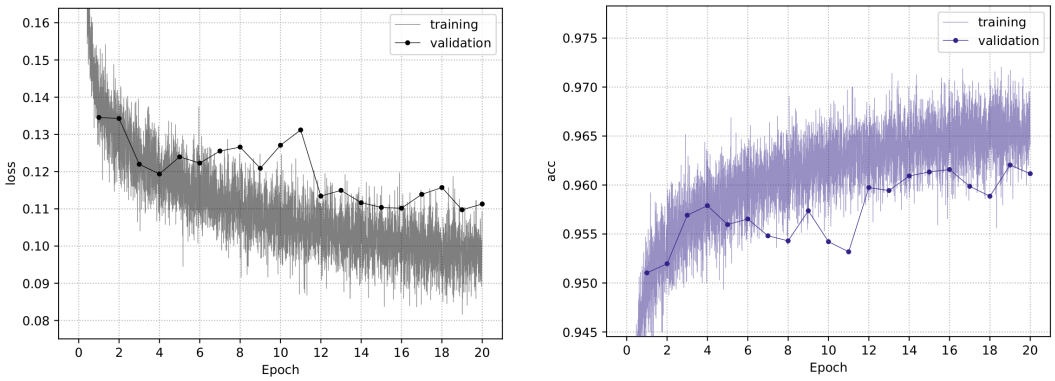


Figure 4.11: ARCA8, k-NN=16 training. The loss curves for the training and validation (with dots) sample for ARCA8 (left) and The accuracy curves for the training and validation (with dots) sample for ARCA8 (right).

are reported. A good agreement between real data and MC simulation is shown in Figures 4.13, 4.15, where the background events are suppressed. A peak of events with very high neutrino score in the data can be observed, which is compatible with an excess of atmospheric neutrinos in that region of the neutrino score. Table 4.5 showcases the evaluation outcomes of the ARCA8 trained network with k-NN=10, utilizing ARCA7 and ARCA8 data. Analyzing the evaluation conducted with ARCA7 data, it is evident that at high neutrino scores (nu_score>0.98), 93.20% and 97.14% of atmospheric and cosmic neutrinos remain respectively. Similarly, the evaluation conducted with ARCA8 data shows that 91.61% and 97.70% of atmospheric and cosmic neutrinos are classified as neutrinos. Also, it is worth noting that 9.52% and 8.18% of atmospheric muons survive the high neutrino score selection in the respective evaluation. Comparing the two inferences, ARCA8 showcases slightly better results than the ARCA7 when using the ARCA8 trained network. Overall, the ARCA8-trained network with k-NN=10 seems to be able to differentiate neutrinos from the background effectively.

In Table 4.6 the evaluation outcomes of the ARCA8 trained network with k-NN=16, using ARCA8 data, are reported. In the training datasets both shower and track neutrinos were included as well as for the network evaluation. After applying the predefined selections ($n\_trig\_hits > 25$, $log\_lik > 50$, $zenith > 90$) and at neutrino scores above 0.98, the contamination from atmospheric muons is 4.33%. Track neutrinos occupy a percentage of 93.18%, whereas shower neutrinos 73.86%. Additionally, 97.36% of cosmic neutrinos remain. The network demonstrates a stronger capability in classifying tracks compared to showers, yet it still performs well in overall classification.

A small difference at the muons-neutrinos events with no selections is observed while looking at Table 4.3 where the ARCA7 data have been evaluated by comparing them with the other ones. This is because all the available runs were used to evaluate the ARCA6 & ARCA8 trainings, whereas for the ARCA7 training some of the events (500K) were used for the training session and the rest for the evaluation.

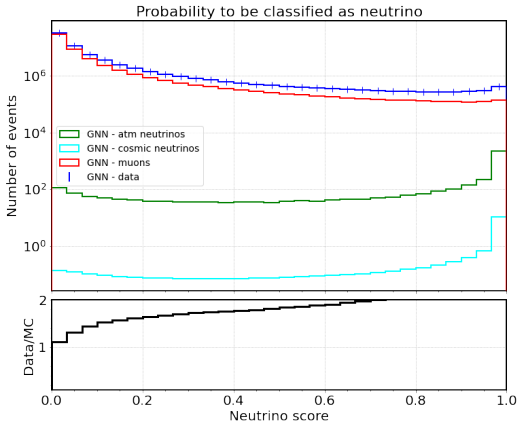


Figure 4.12: ARCA8, k-NN=10 trained network. Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio.

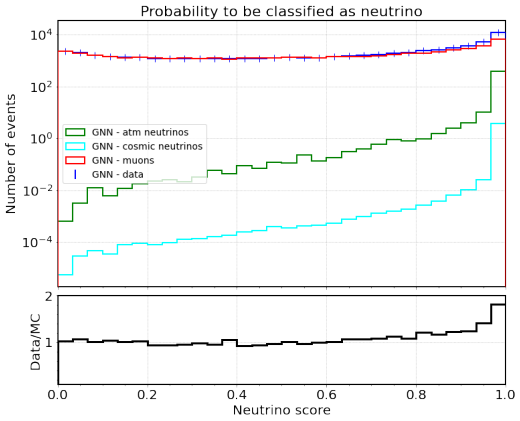


Figure 4.13: ARCA8, k-NN=10 trained network. Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio. Selection criteria: $n\_trig\_hits > 25$, $log\_lik > 50$, $zenith > 90$ (up-going).

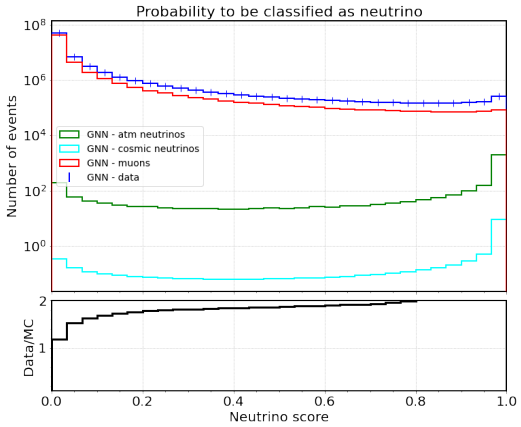


Figure 4.14: ARCA8, k-NN=16 trained network. Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio.

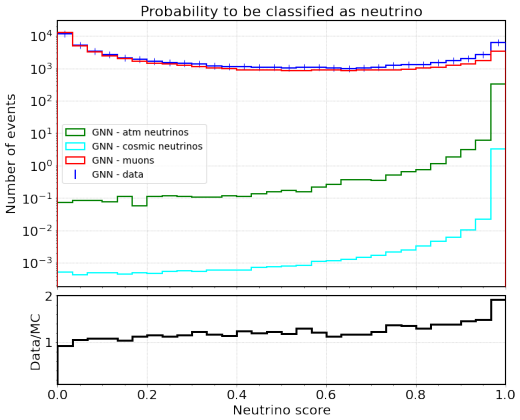


Figure 4.15: ARCA8, k-NN=16 trained network. Upper part: Probability of the event to be classified as neutrino. Lower part: data/MC ratio. Selection criteria: $n\_trig\_hits > 25$, $log\_lik > 50$, $zenith > 90$ (up-going).

ARCA8 TRAINING - ARCA7 INFERRED			
	events w\o cuts	nu score<=0.2	nu score>=0.98
data	11759482 (100%)	9808174 (83.41%)	53572 (0.004%)
atm. muons	9624273.2 (100%)	8437373.6 (87.67%)	18684.6 (0.20%)
atm. neutrino	655.2 (100%)	60.4 (9.22%)	361.3 (55.14%)
cosmic neutrino	2.56 (100%)	0.10 (3.91%)	1.83 (71.48%)
trig. hits>25 & log. lik.>50 & zenith>90			
data	12344 (100%)	2270 (18.39%)	1670 (13.53%)
atm. muons	11273 (100%)	2091.0 (18.55%)	1073.3 (**9.52%**)
atm. neutrino	73.0 (100%)	0.008 (0.01%)	68.0 (**93.20%**)
cosmic neutrino	0.7 (100%)	0.00 (0.006%)	0.68 (**97.14%**)

ARCA8 TRAINING - ARCA8 INFERRED			
	events w\o cuts	nu score<=0.2	nu score>=0.98
data	69414231 (100%)	57313147 (82.57%)	284658 (0.004%)
atm. muons	53307525 (100%)	46768367.3 (87.73%)	84840.66 (0.16%)
atm. neutrino	3958.79 (100%)	379.66 (9.59%)	2068.98 (52.26%)
cosmic neutrino	14.64 (100%)	0.61 (4.17%)	10.26 (70.08%)
trig. hits>25 & log. lik.>50 & zenith>90			
data	67494 (100%)	11257 (16.68%)	9151 (13.56%)
atm. muons	57306.10 (100%)	10601.28 (18.50%)	4687.04 (**8.18%**)
atm. neutrino	421.34 (100%)	0.06 (0.01%)	386.00 (**91.61%**)
cosmic neutrino	3.92 (100%)	0.0003 (0.0001%)	3.83 (**97.70%**)

Table 4.5: Left: Evaluation of the ARCA8, k-NN=10 trained network using ARCA7 data. Right: Evaluation of the ARCA8, k-NN=10 trained network using ARCA8 data.

ARCA8 TRAINING - ARCA8 INFERRED			
	events w\o cuts	nu score<=0.2	nu score>=0.98
data	69414231 (100%)	63057165 (90.84%)	174560 (0.002%)
atm. muons	53307525 (100%)	49972494.2 (93.74%)	53394.4 (0.0001%)
atm. neutrino tracks	3277.69 (100%)	378.34 (11.54%)	1905.95 (58.15%)
atm. neutrino showers	9158.76 (100%)	81.64 (0.001%)	3061.49 (33.43%)
cosmic neutrino	12.6 (100%)	0.88 (6.98%)	8.75 (69.44%)
trig. hits>25 & log. lik.>50 & zenith>90			
data	67494 (100%)	28873 (42.78%)	5037 (7.46%)
atm. muons	57306.1 (100%)	28263.29 (49.32%)	2482.2 (**4.33%**)
atm. neutrino tracks	360.97 (100%)	0.55 (0.002%)	336.35 (**93.18%**)
atm. neutrino showers	1133.44 (100%)	0.04 (0.0001%)	837.14 (**73.86%**)
cosmic neutrino	3.41 (100%)	0.003 (0.01%)	3.32 (**97.36%**)

Table 4.6: Evaluation of the ARCA8, k-NN=16 trained network using ARCA8 data.

4.1.4. GNN performance on ARCA6-8

The comprehensive outcomes are showcased for the three training and evaluation sessions carried out with the ARCA6, ARCA7, and ARCA8 datasets, as reported in the proceedings of the ICRC2023 conference [33]. The inference of the network trained on ARCA6 has been performed on a total lifetime of 45 days, for the ARCA7 trained GNN, a period of 25.5 days has been used, while for the network trained on ARCA8, a period of 22.2 days was examined. In total 93 days have been analyzed. The analysis results are shown in Figure 4.16, where the probability for each event to be classified as neutrino is reported.

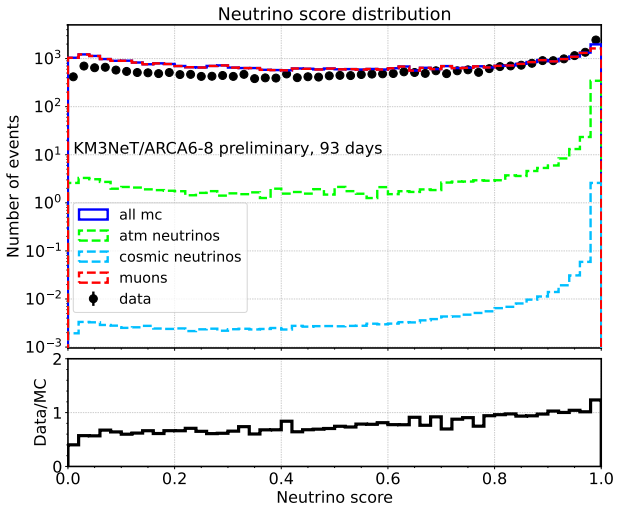


Figure 4.16: Probability of the events to be classified as neutrino for ARCA6-8. Graph taken from [33].

A peak of events with a very high neutrino score in the data, compatible with an excess of atmospheric neutrinos in that region of the neutrino score is observed. The data-Monte Carlo comparison is compatible with values obtained in other KM3NeT analyses, exploiting different selection methodologies [34].

4.2. Track/Shower classification

	track-shower classifier
track neutrinos	379 (40.6%) / 32 (34.0%)
shower neutrinos	470 (50.3%) / 47 (50.0%)
atm. muons	85 (9.1%) / 15 (16.0%)
total	934 / 74

Table 4.7: Number of events used in training and validation datasets for the track/shower classification network using ARCA8 data. The first value represents the absolute number of events in thousands in the training set and the second value the number in the validation set, with the proportion indicated in parentheses.

A classification model has been defined for particle identification. It is trained to learn to characterize the events as tracks or showers. The shower-like neutrinos that are used for the training and evaluation processes are $\nu_\mu NC$, $\overline{\nu}_\mu NC$, $\nu_e CC$, $\nu_e NC$, $\overline{\nu}_e NC$, and the track-like neutrinos are $\nu_\mu CC$, $\overline{\nu}_\mu CC$. The classifier produces a track probability for each event, ranging from 0 to 1. The track-like muons and neutrinos are assigned a value close to 1, while shower-like neutrinos tend to 0. Similarly to the other processes, 90% of the dataset is used for the training and 10% for the validation. The training was performed with about 1 million events of ARCA8 v8.0 MC, equally divided among tracks and showers (see Table 4.7). The network is trained with ReLU as activation function, Adam as optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 0.1$), epochs equal to 20, initial learning rate 0.003 with a 0.0025% decrease after each epoch, batchsize 32 and k-nearest neighbors (k-NN) equal to 16. The model that is finally used to evaluate the network is the one with epoch equal to 19. The evaluation of the network involves analyzing a period of 170 days. The training and validation loss and accuracy are visualized in Figure 4.17. It is noticeable that the validation loss steadily decreases over the epochs reaching lower values compared to training.

In the following plots 4.18, 4.19 the distributions of track scores with weighted events are presented. In Figure 4.18 no selections are applied to the events, whereas in Figure 4.19 a selection on the number of triggered hits for each event $n\_trig\_hits > 25$ and on the log-likelihood $log\_lik > 50$ is applied to suppress the background. In both Figures, the data is denoted by the black line. Muons, shown in red, appear to closely follow the distribution of the data. Track neutrinos are represented by the yellow line and shower neutrinos by the blue line. Upon applying the necessary selections, $n\_trig\_hits > 25$, $log\_lik > 50$, at high track scores a peak of track events is noticeable while the shower events decline. Table 4.8 showcases the evaluation outcomes of the ARCA8 trained network using ARCA8 data. Atmospheric muons are accurately identified as tracks, where 92.63% of them remain when applying track_score > 0.98 and the selections $n\_trig\_hits > 25$, $log\_lik > 50$. A percentage of 35.86% of track neutrinos remain and 0.0001% of shower neutrinos survive the

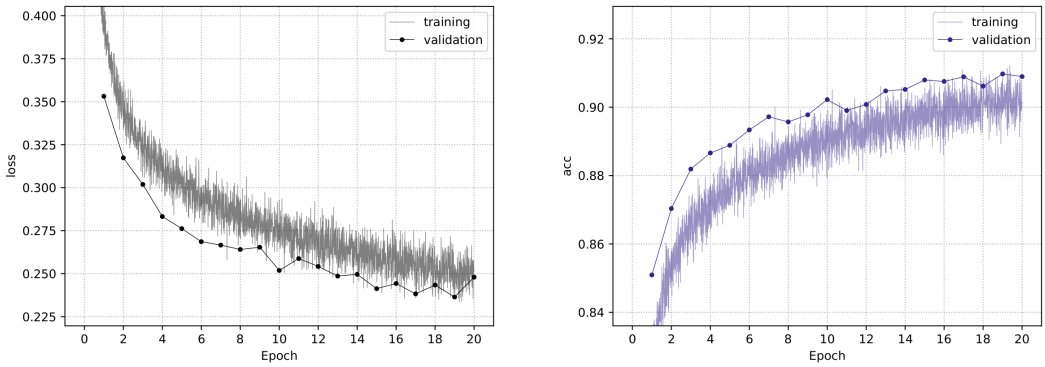


Figure 4.17: ARCA8 training. Left: The loss curves for the training and validation (with dots) sample for ARCA8. Right: The accuracy curves for the training and validation (with dots) sample for ARCA8.

selection. It appears that the network struggles to distinguish track neutrinos from showers, yet correctly assigns a track score closer to 0 for shower events. This is probably due to the fact that most track neutrinos leave only a small percentage of light inside the limited instrumented volume of ARCA8, thus not allowing the discrimination of short tracks and shower topologies.

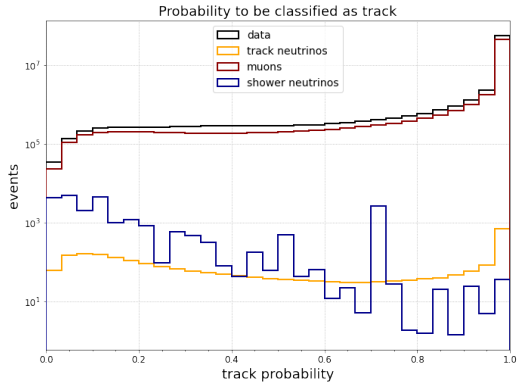


Figure 4.18: ARCA8 trained network. Probability of the event to be classified as track.

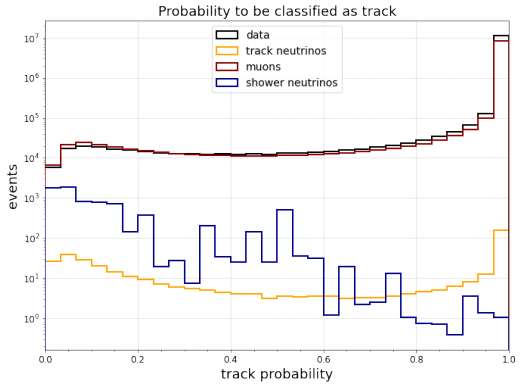


Figure 4.19: ARCA8 trained network. Probability of the event to be classified as track. Selection criteria: $n\_trig\_hits > 25$, $log\_lik > 50$.

ARCA8 TRAINING - ARCA8 INFERRED			
	events w\o cuts	track score<=0.2	track score>=0.98
data	69414231 (100%)	1166906 (1.68%)	54816423 (78.97%)
atm. muons	53307525 (100%)	912259.9 (1.71%)	42442862.8 (79.62%)
track neutrino	2587.68 (100%)	783.00 (30.26%)	647.62 (25.03%)
shower neutrino	24313.02 (100%)	18154.3 (74.67%)	36.21 (0.001%)
trig. hits>25 & log. lik.>50			
data	12918432 (100%)	98928 (0.008%)	12111409 (93.75%)
atm. muons	9520318.6 (100%)	116349.6 (1.22%)	8818445.8 (**92.63%**)
track neutrino	429.4 (100%)	145.02 (33.77%)	153.98 (**35.86%**)
shower neutrino	7679.6 (100%)	6210.02 (80.86%)	0.71 (**0.0001%**)

Table 4.8: Evaluation of the ARCA8 trained network using ARCA8 data.

4.3. Energy regression

In a first attempt to reconstruct the neutrino energy for ARCA8, the GNN is employed. A regression model is defined to produce an estimation of the neutrino energy from the analysis of the signal produced by the propagating particle(s). Along with the predicted value, the algorithm provides an estimated uncertainty. As previously noted, the true label of the energy is the decadic logarithm, $log_{10}E$. This implies that the uncertainty is a factor rather than an absolute value. To illustrate this, consider the following example: The predicted value as generated form the output layer of the network shall be $log_{10}E = 2$ and its uncertainty $log_{10}E_{unc} = 0.1$. This means that the predicted energy is $E = 10^2$ GeV and the boundaries for the uncertainty are calculated as $E_{unc} = 10^{2\pm0.1}$. Then, the upper error is $10^{2+0.1} = 10^{2.1} \approx 125.9$ GeV and the lower error $10^{2-0.1} = 10^{1.9} \approx 79.4$ GeV. It is evident that the uncertainty boundaries are asymmetrical and the uncertainty is the relative factor $10^{0.1} \approx 1.259$.

The GNN for the training phase uses as input the graphs generated from the MC simulation. Similarly to the previous sections, the training process uses approximately 90% of the dataset for training and 10% for validation. In particular, the training was performed with about 1 million events of ARCA8 v8.0 MC with an energy spectrum that ranges from 10^2 to 10^8 GeV, consisting of track-like and shower-like neutrinos, as shown in Table 4.9. The validation dataset was composed of 100k events, in similar proportions. The energy estimation is performed by means of the last fully-connected layer with a linear activation function, which calculates the estimated energy. The network is trained using SGD as optimizer (momentum=0.9, decay=0), epochs equal to 20, initial learning rate 0.009 with a 0.02% decrease after each epoch, batchsize 64 and k-nearest neighbors (k-NN) equal to 16. The best validation epoch turned out to be the 20th. The inference of the network has been performed on a total lifetime of 170 days.

Track-like ((a)numuCC) and shower-like (numuNC, (a)nueCC, anueNC) event topologies are characterized by a different spatial evolution inside the detector, hence in the following figures, the performances are reported separately.

	energy reconstruction
track neutrinos	148 (36.1%) / 32 (36.7%)
shower neutrinos	262 (63.9%) / 55 (63.2%)
total	410 / 87

Table 4.9: Number of events used in training and validation datasets for the energy reconstruction network using ARCA8 data. The first value represents the absolute number of events in thousands in the training set and the second value the number in the validation set, with the proportion indicated in parentheses.

4.3.1. GNN performance on ARCA8

Figure 4.20 shows the training and validation loss curves. The validation loss decreases steadily as the training progresses. It is fair to assume that the model generalizes well to unseen data.

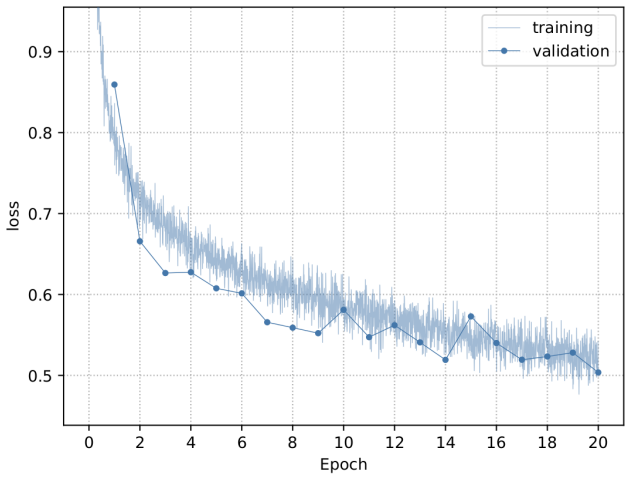


Figure 4.20: ARCA8 training. The loss curves for the training and validation (with dots) sample for ARCA8.

In Figures 4.21, 4.22 the predicted energies with respect to the Monte Carlo energies for track-like and shower-like neutrinos are shown. Most of the events are grouped around the identity line (diagonal) for both shower and track neutrinos. Observing the behavior of track-like neutrinos in Figure 4.21, for low energies up to 10^3 GeV most entries are further away from the diagonal showing a wider data dispersion. As the energy increases they are getting closer to the diagonal showcasing a better reconstruction. A strong selection on the uncertainty ($E_{unc} \leq 1$) reduces the outliers and brings the median closer to the diagonal, successfully eliminating most of the underestimation. In the case of shower-like events, Figure 4.22, the reconstructed energy for the majority of the events is already close to the diagonal showcasing a narrow data dispersion. Applying the same uncertainty selection small changes are observed where mostly the lower-energetic events are affected. It is worth noticing the better performances of the shower-like events with respect to track-like ones. This behavior is probably due to the better event containment for showers.

Similarly , the graphs 4.23 and 4.24 show the predicted energies with respect to the Monte Carlo energies for shower-like and track-like events respectively. The coefficient of determination, R^2 score, is computed to assess the degree of alignment between the predicted and true energies, using the formula $R^2 = 1 - \frac{SSR}{SST}$, where SSR is the sum of squared residuals $\sum_{i=1}^{n}(y_i^{true} - y_i^{pred})^2$ and SST is the total sum of squares $\sum_{i=1}^{n}(y_i^{true} - \bar{y}^{true})^2$. For shower-like and track-like events R^2 is 0.708 and 0.304 respectively.

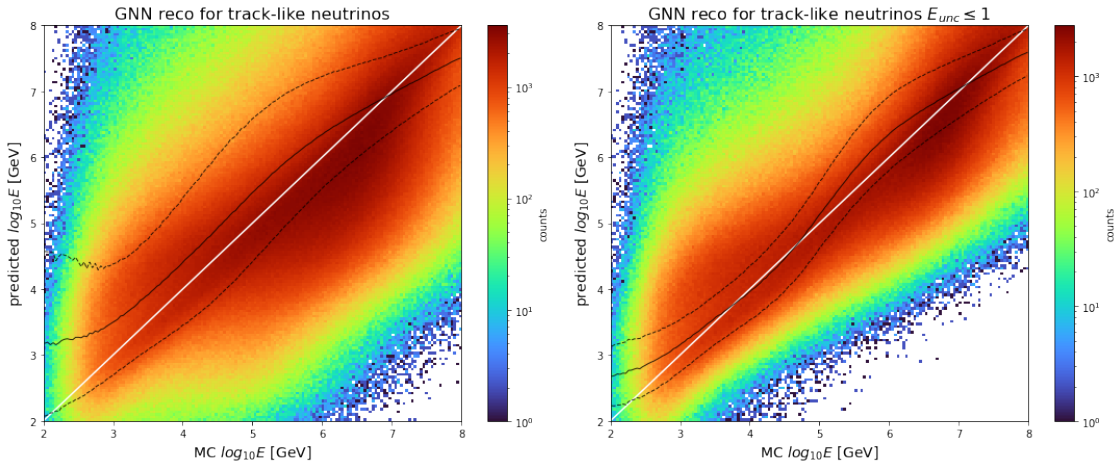


Figure 4.21: ARCA8 trained network. Reconstructed versus true energy for track neutrinos. Left: No selection is applied. Right: Uncertainty selection ($E_{unc} \leq 1$). A solid white line is added as a diagonal to indicate the perfect reconstruction. The black solid line is the median and the black dashed lines represent the uncertainties of the data.

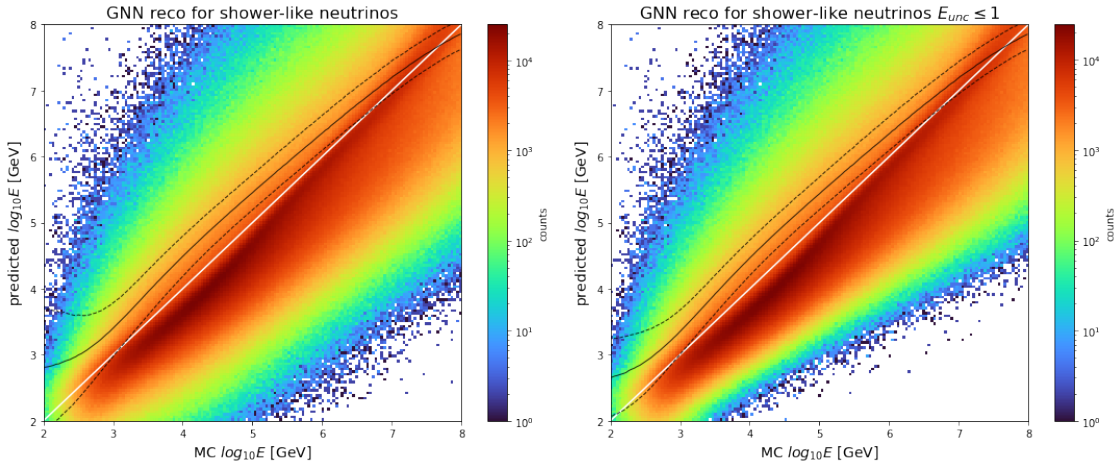


Figure 4.22: ARCA8 trained network. Reconstructed versus true energy for neutrino interactions creating shower signatures. Left: No selection is applied. Right: Uncertainty selection ($E_{unc} \leq 1$). A solid white line is added as a diagonal to indicate the perfect reconstruction. The black solid line is the median and the black dashed lines represent the uncertainties of the data.

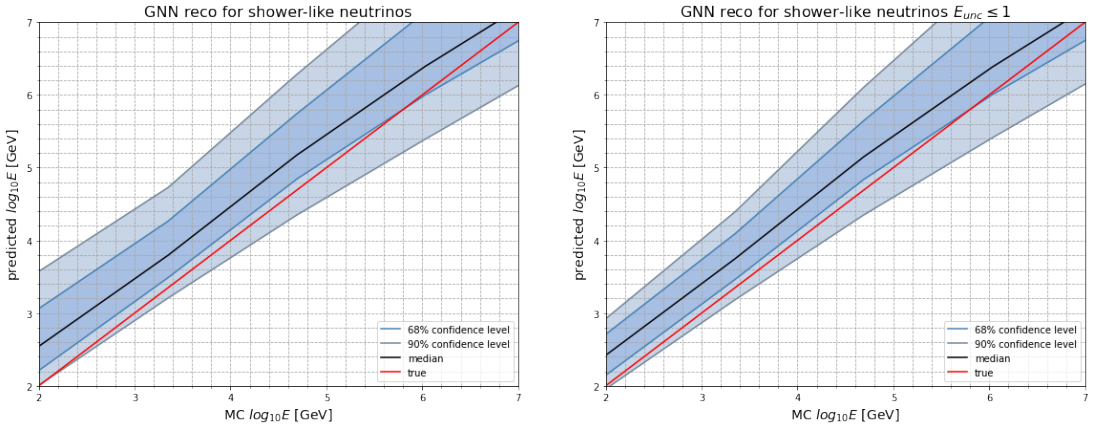


Figure 4.23: ARCA8 trained network. Reconstructed versus true energy for neutrino interactions creating shower signatures. Left: No selections applied. Right: Uncertainty selection ($E_{unc} \leq 1$). The confidence intervals at 68% and 90% are illustrated.

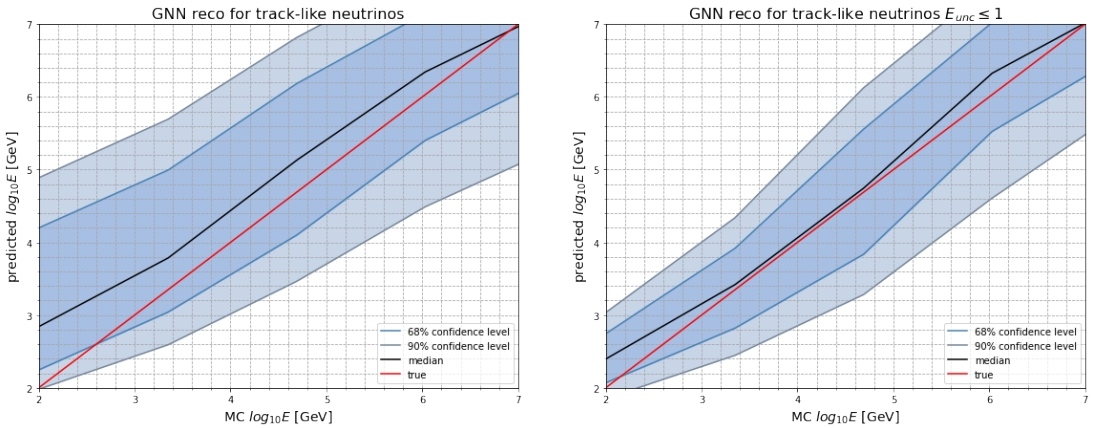


Figure 4.24: ARCA8 trained network. Reconstructed versus true energy for neutrino interactions creating shower signatures. Left: No selections applied. Right: Uncertainty selection ($E_{unc} \leq 1$). The confidence intervals at 68% and 90% are illustrated.

5
Conclusion

In Chapter 1 the fundamental properties of neutrinos as well as its detection techniques are presented.

Chapter 2 analyses the KM3NeT research infrastructure consisting of two neutrino detectors KM3NeT/ARCA and KM3NeT/ORCA. A brief technical description of the detectors is provided.

Chapter 3 introduces a subfield of Deep Learning, Graph Neural Networks (GNNs). The Graph Neural Networks are trained on Monte Carlo (MC) simulations of events with the KM3NeT/ARCA detector. While the full detector will consist of 115 so-called detection units, this thesis uses simulations from an early stage of the detector consisting of six, seven and eight detection units.

Chapter 4 demonstrates the performance of the GNN for three different networks: signal/background classifier, track/shower classifier, and energy regression.

Firstly, the signal/background classification is examined using KM3NeT/ARCA data collected from 6, 7, and 8 DUs. The KM3NeT/ARCA6 trained network is evaluated using ARCA7 and ARCA8 data separately. However, the classifier encounters challenges in effectively distinguishing between neutrinos and the background. Subsequently, the KM3NeT/ARCA7 network is trained and evaluated. The classifier effectively classifies neutrinos and atmospheric muons, demonstrating slightly better classification when applied to ARCA7 data compared to ARCA8. Following this, two KM3NeT/ARCA8 networks are trained and evaluated, differing mainly in the k-NN parameter set to 10 and 16, respectively. Notably, the KM3NeT/ARCA8 with k-NN=16 exhibits a slightly better performance in recognizing neutrinos compared to KM3NeT/ARCA8 with k-NN=10.

Moving on, the track/shower classification is assessed using KM3NeT/ARCA8 data. The network faces challenges in distinguishing track neutrinos from showers, but it accurately assigns a track score closer to 0 for shower events.

Finally, the energy regression is attempted using KM3NeT/ARCA8 data, revealing better performance for shower-like events compared to track-like ones. This behavior is likely attributed to the better event containment for showers.

References

[1] Qu Huilin ; Gouskos Loukas. "ParticleNet: Jet Tagging via Particle Clouds". In: *Phys. Rev. D* 101.5 (2020), p. 056019. DOI: 10.1103/PhysRevD.101.056019. arXiv: 1902.08570 [hep-ph].

[2] Mary K. Gaillard ; Paul D. Grannis ; Frank J. Sciulli. "The standard model of particle physics". In: *Reviews of Modern Physics* 71.2 (1999).

[3] Paolo Lipari. "INTRODUCTION TO NEUTRINO PHYSICS". In: (2003). https://cds.cern.ch/record/677618/files/p115.pdf.

[4] A.Y. Smirnov. "THE MIKHEYEV-SMIRNOV-WOLFENSTEIN (MSW) EFFECT". In: *arXiv* (Feb. 2017). https://arxiv.org/pdf/1901.11473.pdf.

[5] Kohta Murase. *Active Galactic Nuclei as High-Energy Neutrino Sources*. 2017.

[6] R. N. Mohapatra ; Palash B. Pal. *Massive neutrinos in Physics and Astrophysics*. World Scientific Publishing Co. Pte. Ltd., 2004. ISBN: 981-238-070-1.

[7] Daniel Guderian. "Development of detector calibration and graph neural network-based selection and reconstruction algorithms for the measurement of oscillation parameters with KM3NeT/ORCA". https://www.uni-muenster.de/imperia/md/content/physik_kp/agkappes/abschlussarbeiten/doktorarbeiten/doktorarbeit_daniel_guderian.pdf. PhD thesis. Mathematisch-Naturwissenschaftlichen Fakultät der Westfälischen Wilhelms-Universität Münster, 2022.

[8] S. Adrián-Martínez et al. "Letter of intent for KM3NeT 2.0". In: *J. of Physics G: Nuclear and Particle Physics* 43.8 (June 2016), p. 084001. DOI: 10.1088/0954-3899/43/8/084001.

[9] Tommaso Chiarusia and Paolo Piattelli. "KM3NeT: R&D and technical solutions for the next generation underwater neutrino telescope". In: *Nuclear and Particle Physics Proceedings* 273-275 (2016), pp. 2357–2359. DOI: https://doi.org/10.1016/j.nuclphysbps.2015.09.390.

[10] S. Aiello ; F. Ameli ; M. Andre et al. "The Control Unit of the KM3NeT Data Acquisition System". In: *Computer Physics Communications 256* (2020). DOI: https://doi.org/10.1016/j.cpc.2020.107433.

[11] Angela Zegarelli. "Exploring the high-energy transient sky with ANTARES and KM3NeT neutrino telescopes". PhD thesis. Tor Vergata University, Rome, Italy Sapienza University, Rome, Italy, 2023.

[12] K. Melis ; A. Heijboer ; M. de Jong. “KM3NeT/ARCA Event Reconstruction Algorithms”. In: *PoS* ICRC2017 (2017). https://pos.sissa.it/301/950/pdf, p. 950.

[13] B. O. Fearraigh ; Bouke Jung. “Track Reconstruction in KM3NeT”. In: *internal note* (2021). https://common.pages.km3net.de/jpp/TrackReconstruction.PDF.

[14] F. James ; M. Winkler. *MINUIT User‘s Guide: C++ Version*. 2004.

[15] ANTARES Collaboration. “ANTARES: the first undersea neutrino telescope”. In: *Nuclear Instruments & Methods in Physics Research Section A* (2011). DOI: https://doi.org/10.1016/j.nima.2011.06.103.

[16] H. Hanada et al. “A highly sensitive optical detector for use in deep underwater”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 408 (1998), pp. 425–437. DOI: https://doi.org/10.1016/S0168-9002(98)00196-X.

[17] AMANDA collaboration. “The AMANDA neutrino telescope: principle of operation and first results”. In: *Astroparticle Physics* 13 (2000), pp. 1–20. DOI: https://doi.org/10.1016/S0927-6505(99)00092-4.

[18] IceCube Collaboration. “The IceCube Neutrino Observatory: Instrumentation and Online Systems”. In: *JINST 12 P03012* (2017). DOI: https://doi.org/10.48550/arXiv.1612.05093.

[19] IceCube Collaboration. “IceCube-Gen2 : A Vision for the Future of Neutrino Astronomy in Antarctica”. In: *arXiv e-prints* (Dec. 2014). DOI: 10.48550/arXiv.1412.5106.

[20] Kenneth Greisen. “End to the cosmic-ray spectrum?” In: *Physical Review Letters* 16.17 (Apr. 1966), pp. 748–750.

[21] Zatsepin G. T. ; Kuz’min V. A. “Upper Limit of the Spectrum of Cosmic Rays”. In: *Journal of Experimental and Theoretical Physics Letters* 4 (Aug. 1966), pp. 78–80.

[22] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[23] M. A. Nielsen. *Neural Networks and Deep Learning*. https://www.ise.ncsu.edu/fuzzy-neural/wp-content/uploads/sites/9/2022/08/neuralnetworksanddeeplearning.pdf. Determination Press, 2015.

[24] Diederik P. Kingma ; Jimmy Lei Ba. “Adam: A Method for Stochastic Optimization”. In: *arXiv* (2017). https://arxiv.org/pdf/1412.6980.pdf.

[25] Andriy Burkov. *The Hundred-Page Machine Learning Book*. 2019.

[26] P. W. Battaglia et al. “Relational inductive biases, deep learning, and graph networks”. In: *arXiv e-prints* (2018). https://arxiv.org/pdf/1806.01261.pdf.

[27] J. Shlomi ; P. Battaglia ; J.-R. Vlimant. “Graph neural networks in particle physics”. In: *Machine Learning: Science and Technology 2* 021001 (2021). DOI: http://dx.doi.org/10.1088/2632-2153/abbf9a.

[28] Wang Y. et al. “Dynamic Graph CNN for Learning on Point Clouds”. In: *ACM Transactions on Graphics* 38.5 (Oct. 2019). DOI: https://doi.org/10.1145/3326362.

[29] KM3NeT collaboration. https://ml.pages.km3net.de/OrcaSong/.

[30] KM3NeT collaboration. https://km3py.pages.km3net.de/km3pipe/.

[31] KM3NeT collaboration. https://ml.pages.km3net.de/OrcaNet/.

[32] Lucas Hennig. *Tau neutrino identification with Graph Neural Networks in KM3NeT/ORCA*. 2023.

[33] F. Filippini; E. Androutsou; A. Domi; B. Spisso; E. Drakopoulou. “Data reconstruction and classification with graph neural networks in KM3NeT/ARCA6-8”. In: *PoS* ICRC2023 (2023). https://pos.sissa.it/444/1194/pdf, p. 1194.

[34] V. Tsourapis et al. “Search for a diffuse astrophysical neutrino flux using ARCA data”. In: *PoS* ICRC2023 (2023). https://pos.sissa.it/444/1195/pdf, p. 1195.