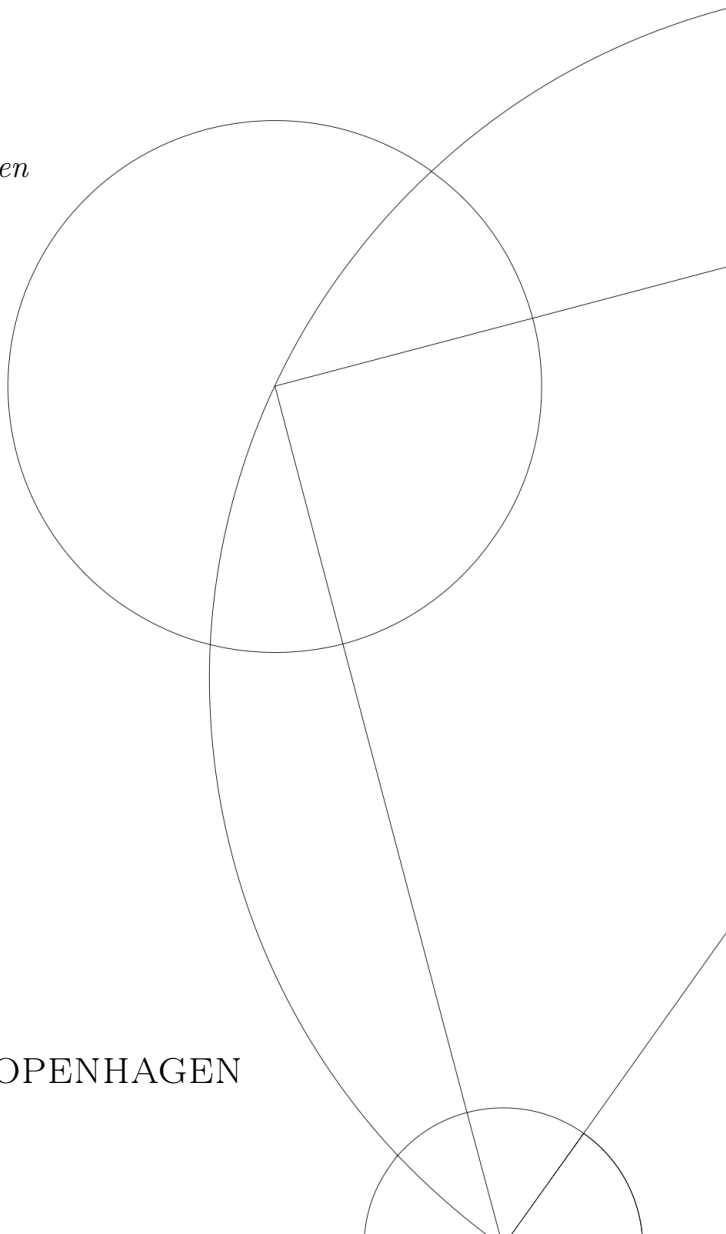
Identifying and modeling stopped muons using temporal convolutional networks on IceCube data

Bachelor project

Written by *Simon C. Debes & Frederik V.S.S. Hansen*

November 20, 2021

Supervised by
Troels C. Petersen



University of Copenhagen

UNIVERSITY OF COPENHAGEN

Name of institute:	Københavns universitet
Name of department:	Niels Bohr Institutet
Authors:	Simon C. Debes & Frederik V.S.S. Hansen
Email:	rlq306@alumni.ku.dk & cjb924@alumni.ku.dk
Title and subtitle:	Identifying and modeling stopped muons using temporal convolutional networks on IceCube data -
Supervisor:	Troels C. Petersen
Handed in:	16/06-2021
Defended:	25/06-2021

Name _____

Signature _____

Date _____

Abstract

This project investigates the effectiveness of temporal convolutional neural networks in reconstruction of simulated muon events at the IceCube neutrino observatory. There has been produced four models: A classifier to identify whether a muon had stopped in the detector or passed through. This model achieves an AUC score of 0.885, but probably has a higher potential. Surprisingly there appeared to be no correlation between the classifiers accuracy, and the energy of the muon. There have also been made three regression models for energy, zenith angle and azimuth angle respectively. The energy regression model produced predictions that compared to the target values with $\sigma = 0.144$ GeV. The zenith regression model was the most accurate, with a $\sigma = 0.166$ radians, while the azimuth regression was the poorest, with $\sigma = 1.396$ radians. On average, one prediction for each of the four models took $1.5 \cdot 10^{-4}$ s to $2 \cdot 10^{-4}$ s to compute on a NVidia GeForce RTX 2070 GPU.

Acknowledgements

Without the help of many people this thesis would never have looked as it does today. First of all we need to thank our supervisor Troels Petersen for taking us in as bachelor students and giving us a great learning environment.

We also a owe a huge thank you to Rasmus Ørsøe who has already always been ready to answer any questions and help with debugging code.

At last we would like to thank the rest of the people from our small IceCube machine learning group for always being helpful when needed.

Contents

1 Reader's guide

This section is a reading guide which provides a context to the reader from which the work done in this thesis can be understood.

1.1 The Problem

The motivation behind this thesis comes from the ultimate goal of physics: To find the theory of everything. The closest thing we have today is the standard model, which we have found to be flawed, including its description of neutrinos. Neutrinos are some of the least explored pieces of the standard model, and therefore also a good place to start if attempting to "complete" the standard model. At the IceCube observatory they are using the better known muons to help reconstruct and understand the neutrinos better. Neutrinos are very rarely detected, making it difficult to model precisely. Furthermore the properties of the ice and the detectors are not very well known, making it even more problematic. Due to their higher interaction rate, muons are detected a lot more frequently, and sometimes stop their trajectory inside the detector, allowing us to more accurately describe their properties. Stopped muons are the only source to signal where it is possible with high statistics to compare simulation with real data and thereby calibrate and understand systematical errors. This is something IceCube has not been able to accomplish yet with high statistics, as their statistical method of reconstruction called **RetroReco** takes 5-40 seconds to make a reconstruction[1]. We believe that it is possible to achieve comparable results, if not better, using modern machine learning techniques, orders of magnitude faster.

I: A binary classifier to determine if a muon traveled through the detector or stopped inside the detector.
II-IV: Regression models to determine the energy of the muon and the zenith and azimuth angle of entry.

1.2 Structure of this work

This paper is organized as follows. In section 2 the reader will be introduced to a short overview of the relevant physical theories necessary for understand what is happening in the IceCube detector. Thereafter section 3 will shortly provide information of what the IceCube detector is, how it is designed and how it measures different particles. Then the reader will be given a short overview of machine learning and how it works in this thesis (section 4). After the first three sections which are more theory based the more discussion based part of the thesis will begin. First the reader will be introduced to the thought process of designing the different models and the reasoning behind the choices of certain parameters in section 5. Section 6 will then present the results of this thesis and discuss how well they are and how they possibly could be improved. In relation with section 6, section 7 will give a conclusion and the last discussion of the results of the thesis. At last section 8 will discuss possible ideas to look further into to improve and continue the work of this thesis.

1.3 Distribution of work

The work done in this thesis[1] has been evenly distributed between the authors Simon Debes and Frederik Hansen. At first both looked at creating a working TCN-model for stopped muon classification and shared the progress between each other. After the stopped muon classifier started working Simon went on to setup the model for the energy regression meanwhile Frederik optimized the stopped muon classifier. Later both worked on the angle regression models and shared the progress between each other.

[1]The written code can be seen here: https://github.com/FredeVHansen/TCN_IceCube_Muon_Reconstruction or here https://github.com/sdebes/icecubeml

2 The underlying physical theory

Even though this project aims to use machine learning which is more computer based work, all the data behind and the results come from physical theories and experiments. Therefore it is important to know the theories behind what we are looking at and how it works, which will be covered in this section.

2.1 Standard Model (SM)

If you have ever wondered why the material world is at it is, like why do flowers smell or why does the sun shine, you will often be able to find a certain answer. But if you keep asking further into the problem you will end up with the Standard Model (SM) as the answer (except if your question involves gravity).[2]

The SM describes how the fundamental particles (figure 1) make up all known matter and how force carriers influence these particles. The SM can be called "scientists' current best theory to describe the most basic building blocks of the universe"[3], where the basic building blocks are the fundamental particles, fermions and bosons, that depend on the four fundamental forces; the strong nuclear force, the weak nuclear force, electromagnetism and the gravitational force. The SM then describes how the fundamental particles and three of these four forces are related (since the gravitational force is not yet unifiable with the SM).

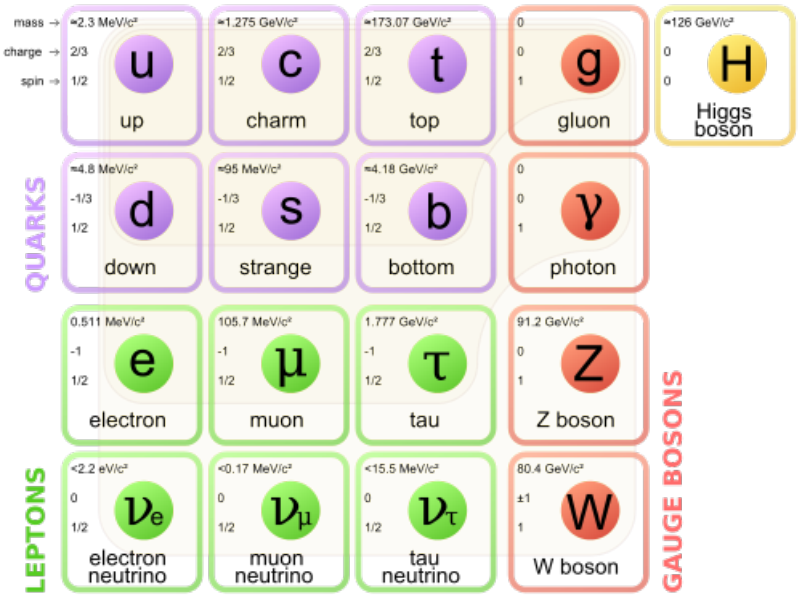


Figure 1: The standard model three generations of particles (first three columns), the force cariers (Gauge bosons in fourth column) and at last the Higgs boson in the fifth column. Image from[4]

As we can see from figure 1 the standard model has three generations of matter which differ in mass. The first generation consists of the up (u) and down (d) quarks and the well known electron particle (e). The second consist of the charm and strange quarks and the muon (μ) particle which we are looking at in this project and is of great interest to IceCube. At last we have the third generation which consist of the top and bottom quarks and the tau particle (τ). Also in each generation there is the corresponding neutrino (ν_e, ν_μ, ν_τ) to each charged lepton (e, μ and τ) where the neutrinos are what IceCube really wants to look at and understand even more.

In this project, we are only concerned with the leptons (and the relevant force carriers), and will not cover the quarks, gluon or Higgs boson. Like the quarks, the leptons come in three

generations. With each generation, the mass is increased, and the lifetime is decreased by many orders of magnitude.

The two forces of the standard model which are of interest to this project, is the electromagnetic force (mediated by the photon), that is responsible for the Cherenkov radiation, which makes particle detection possible, along with the weak force, which is the only force the neutrino can interact through, besides the negligible gravitational interaction. The forces are mediated by the photon, $W^{\pm}$ and Z^0 bosons respectively and the weak force will be discussed in section 2.3 and the muon and neutrino will be discussed in section 2.4 and 2.5.

2.2 Cherenkov radiation

IceCube detects particles using photo multiplier tubes (PMTs), meaning a mechanism allowing the particles to produce light is needed. That mechanism is called Cherenkov radiation.

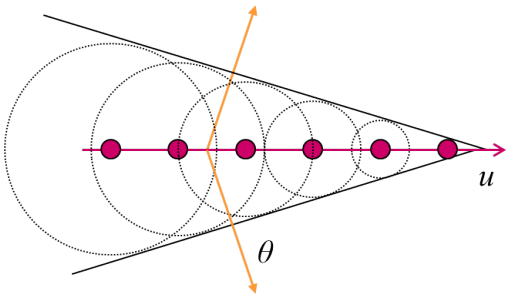


Figure 2: Cherenkov radiation. The muon travels along the pink arrow in some matter, that produces spheres of light with centers in the red dots. The higher the velocity, the narrower the angle of the cone . Image from[5]

It is commonly known that light will travel at a velocity of $c \approx 3 \cdot 10^8\ m/s$ in a vacuum. In dielectric medias though, the refractive index η of the media determines the lights phase velocity, v, through the equation:

$$v = \frac{c}{\eta}. \tag{1}$$

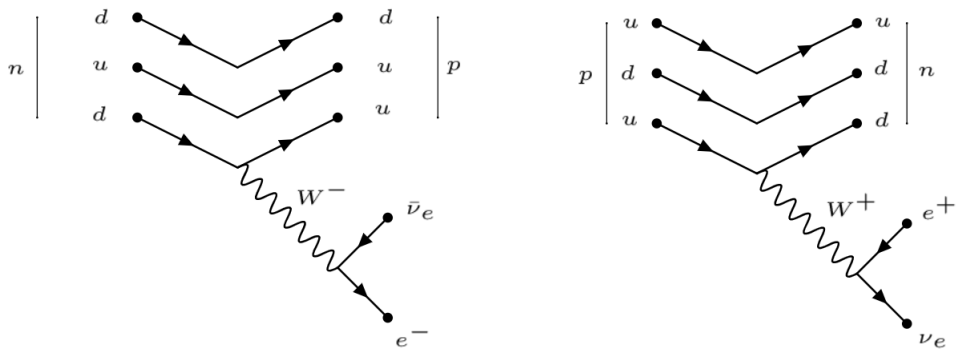
The ice in IceCube has a refractive index $\eta \approx 1.3$ [6], meaning that a particle needs only a velocity of $v \approx \frac{3}{4}c$ to produce Cherenkov radiation.

When a charged muon travels trough ice, it will polarize the particles in the ice, and when travelling away, it will de-polarize the particles which causes the particles to emit photons, which is what we call Cherenkov radiation. Neutrinos posses no charge and are therefore not able to produce Cherenkov radiation. Instead, they can collide with other particles and through weak interaction, other charged particles are created, which then can produce Cherenkov radiation and a neutrino can be observed.

2.3 Weak nuclear force interaction

As mentioned in section 2.1 the weak force is one of the four fundamental forces that govern the matter in the universe.At short ranges, the weak force is comparable to the strong force in strength, but the strength quickly decreases as the range increases, meaning two particles have to be very close to interact weakly, unlike the electromagnetic force which can interact at longer ranges. Since atoms are above 99.99% free space, the neutrino very rarely get close enough to the electrons or nuclei of atoms to interact.

The weak force interaction is mediated by the charged $W^{\pm}$ bosons and the neutral Z^0 bosons. It will change the flavor of a quark by emitting an electrically charged W boson which causes a proton to change into a neutron.

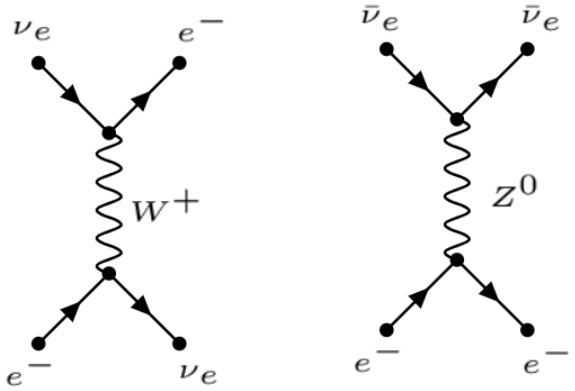


(a) (b)

Figure 3: Figure (a) shows a β^--decay (b) shows a β^+-decay.

The neutrino was first discovered in β-decay (figure 3, a charged current interaction where a neutron and proton exchange a $W\pm$ boson, and emit an $e^{\pm}$ and either a neutrino or anti-neutrino).

We must also remember that the Z-boson is a part of the weak force as it plays a role in neutral current interaction. As we just have seen the $W^{\pm}$ is used in charged currents as the W-boson is a charged particle whereas the Z-boson is a neutral charged particle meaning that interactions in weak force can happen under charged current and neutral current as seen in figure 4. The difference between these two interactions is that in charged current interactions there is produced a charged lepton which gives a signal in the IceCube detector, but for the neutral charge interaction there is produced a neutrino, which they do not see in the IceCube detector.



(a) (b)

Figure 4: Figure (a) shows a Feynman diagram of a current charge interaction between an electron and an electron neutrino (b) shows a neutral charge interaction between an electron and an electron neutrino.

2.4 Muons

Most muons are created when a proton enters the earths atmosphere, and interacts with the nucleus of an atmospheric particle through the strong force and produce a shower of other unstable particles as seen in figure 5. Most often it will be a pion, which decays into a muon and muon neutrino. Other leptons are also produced, but electrons have a very small mass, and cannot travel very far before being stopped. The tauon has a much larger mass, but it's short

lifetime[2] means it also decays very fast. Muons are in the sweet spot regarding lifetime and mass, meaning that they will leave long tracks and can be stopped in a detector just like IceCube.

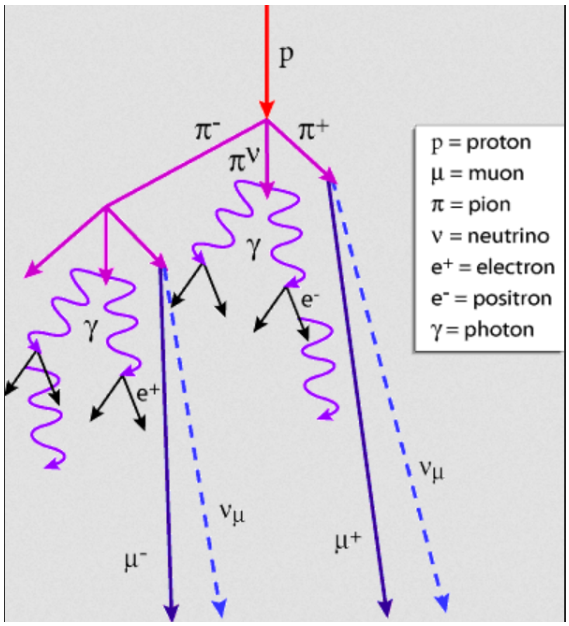


Figure 5: The airshower produced when a proton enters Earths atmosphere. Image from [8]

Muons are charged, fast, and abundant, so they will produce Cherenkov radiation directly, and thus lots of neat data. In fact IceCube detects about 900,000 muon for every neutrino [9], so even though IceCube is a neutrino observatory, the muons are of interest, due to the emphatic knowledge of their movement in the detector. Because we have mapped them so well, the knowledge gained in reconstructing muons, can be transferred to reconstructing the lesser known neutrinos.

[2] $t_\mu \approx 2.2 \cdot 10^{-6}$ s & $t_\tau \approx 3 \cdot 10^{-13}$ s [7]

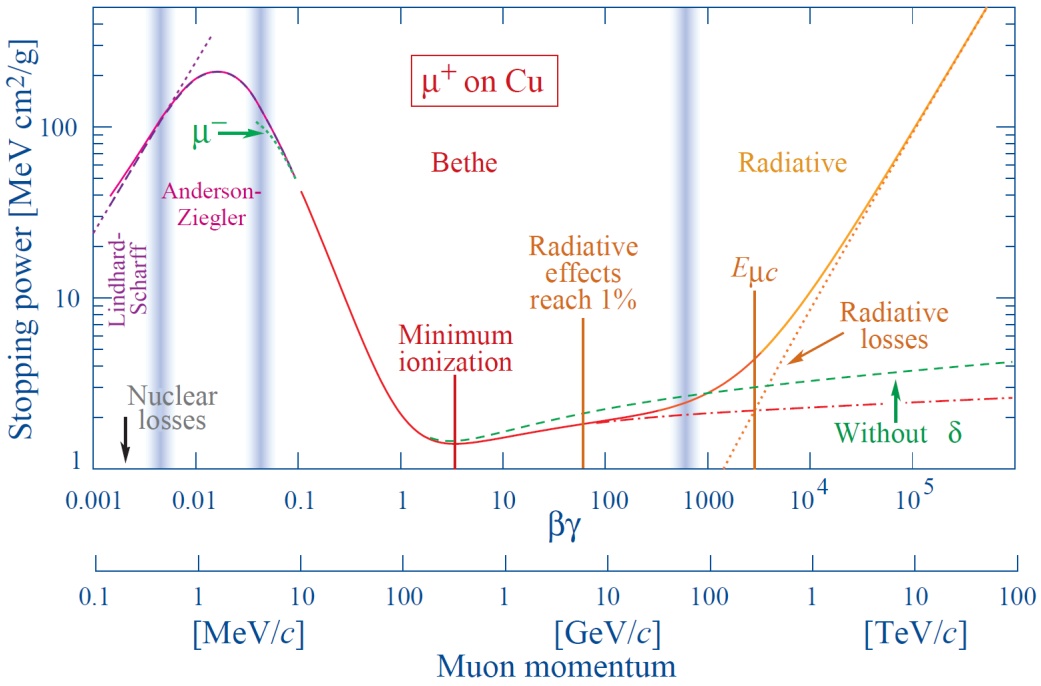


Figure 6: Muon energy loss when traveling in copper (the structure of the plot looks the same for all other materials). The interval between $\approx 0.1GeV$ and $\approx 200GeV$, is where the muon stopping power is the lowest. Image from[10]

Due to the muon's energy loss mechanism (figure 6), muons will lose energy at a high rate until it reaches around 200 GeV. Muons that hit Earths surface with too low energies will not reach the detector as they first have to penetrate 1.400 meters of ice. The shape of the plot in figure 6 is very lucky, because it allows muons of a wider range of energy to stop in the detector. If muons lost energy at a constant rate, stopped muons of certain angles would only have energies in a narrow interval, however due to the high energy loss of high energy muons, this interval is widened.

A 10 TeV muon will lose energy at a high rate until it has 200 GeV, and then lose energy at a slower rate, until it only has 0.1 GeV, where it again will lose its remaining energy quickly, and then stop. Because the muon loses energy at a lower rate in the interval 0.1 GeV to 200 GeV, it is also the interval it spends the most time, and therefore is most likely to be observed. This allows muons of high energies to also be able to stop in a detector like IceCube. If muons lost energy independent of energy, all stopped muons would have energies in a narrow interval.

If one can pinpoint the vertex at which the muon has stopped, and its trajectory, one can integrate figure 6 to find the energy it had when it entered the detector. This is why stopped muons contain more information.

2.5 Neutrinos

Neutrinos, italian for "little neutron", are particles so light, we have a hard time determining their mass. They are leptons (so no strong interaction), they have a Small mass and they got no charge (so no EM interaction). They have only weak interaction and gravity, but we ignore gravity in this thesis. This means that they very rarely interacts. In fact, a beam of neutrinos would need to travel in lead for 1 lightyears before half would have stopped [11]. The fact that the neutrinos has these properties makes them interesting because the SM predicts them to be massless even though they got a small one. And because they are so elusive physicist at IceCube and other particle physicist in general are very curious to know more about them.

2.5.1 Neutrino oscillation

There exists three flavors (or generations) of neutrinos (ν_e, ν_μ, and ν_τ). In the SM, they are assumed to be massless, but as mentioned earlier they have a small mass which is still not well known. Because neutrinos have mass, they can participate in neutrino oscillation[12], where they oscillate between the three different flavors erratically.

These three flavors, $|\nu_\alpha\rangle$ ($\alpha = e, \mu, \tau$), exist in superpositions of the three possible mass states, $|\nu_j\rangle$ ($j = 1, 2, 3$). As the neutrinos propagate, the combination by which they are mixed changes, and is determined by the unitary mixing matrix, $\mathcal{U}$, called the **Pontecorvo-Maki-Nakagawa-Sakata** (PMNS) matrix

$$|\nu_\alpha\rangle = \sum_{j=1}^{3} \mathcal{U}^{\star}_{\alpha j} |\nu_j\rangle \,. \tag{2}$$

The PMNS matrix is parametrized by three rotation matrices, containing the four free parameters, the mixing angle between each of the three pair-permutations, $\phi_{12}, \phi_{13}, \phi_{23}$, and a phase that accounts for charge-parity violation, δ_{CP}. The matrix containing the mixing angle ϕ_{12} can be seen here:

$$\mathcal{U}_{12} = \begin{bmatrix} \cos\phi_{12} & \sin\phi_{12} & 0 \\ -\sin\phi_{12} & \cos\phi_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

This oscillation means that neutrinos are even more complex and difficult to understand, as they are ever-changing.

3 IceCube, the telescope in the ice

Deep below the Amundsen-Scott station on the geographical south pole is a telescope in the ice, observing the cosmos, called the IceCube experiment. IceCube is known as a multipurpose experiment which addresses multiple big questions in physics like the nature of dark matter, the properties of the neutrino itself and many more[13].

Light is how we observe things in our day to day lives, and therefore also the most obvious choice for what humans choose to base the first telescopes on. Since neutrinos interact so rarely compared to photons, they are rarely attenuated, deflected, or stopped, and provide an entirely different picture of the universe. Therefore a telescope that looks at neutrinos is able to observe things that we were not able to observe previously.

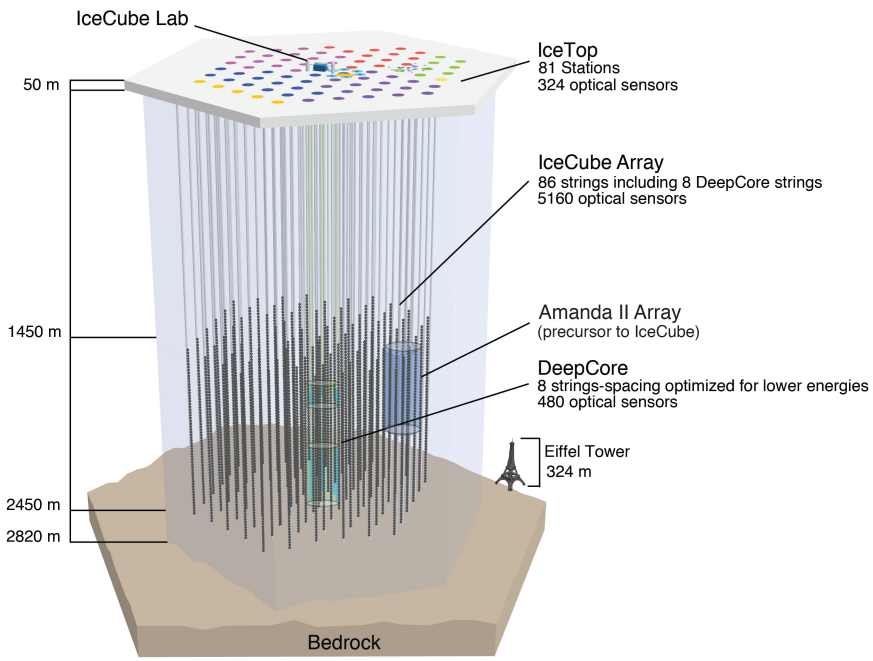
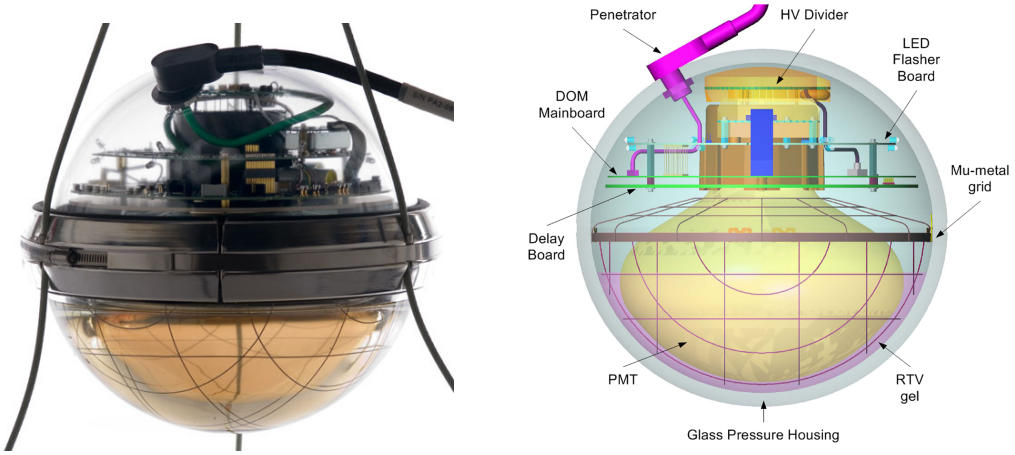


Figure 7: The IceCube deetector with IceTop, IceCube Array, Amanda II Array and DeepCore[14]

3.1 IceCube

At first glance IceCube looks like a building standing on top of the ice but this is just an illusion for the huge experiment which IceCube indeed is. Below the ice top and 2,500 meters down to the antarctic bedrock the real experiment is. 86 strings of Digital Optical Modules (DOM) is spread out beneath the ice 125 meters apart from eachother with 60 DOMs on each string with 17 meters apart. This makes up for a total of 5,160 DOMs in the ice from 1,450 meters beneath the ice top to about 2,500 meters beneath the ice top[15]. Later, a more concentrated cluster of improved DOMs were installed better known as DeepCore which were only installed in the clearest ice 2.100 meters-2450 meters below the surface where 400 DOMs where placed with 7 meters apart vertically. There were no DOMs placed in the region 2.000 meters - 2.100 meters du to a signifanct scattering and absorbtions i that area instead 10 DOMs on each string has been placed 10meters apart vertically directly above this region. All this was done in order to detect neutrinos at lower energies than first envisioned. [16]

3.1.1 Digital Optical Modules (DOMs)



(a) (b)

Figure 8: Figure (a) shows a picture of a DOM while figure Image from[17](b) shows the details of the DOM[18].

Submerged in the ice, we find the glass spheres containing PMTs, where all signals are recorded, called DOMs. When a DOM is triggered, it logs the measurements in a very short timewindow around the measurement.

Due to the gridlike structure of IceCube (as seen in figure 7), there are "corridors" in the detector, where if particles enters at the right angle, they can travel through in a straight line parallel to the DOMs, without ever getting closer than ∼60m to a DOM. Thus the light produced by low energy muons will be attenuated before reaching a DOM, and will not be detected.

A muon and particles in general can also travel through a corridor for a while, and then first be detected well within the detector, with the lack of a Cherenkov trail giving the illusion that it originated inside the ice.

3.1.2 Photomultiplier tubes and noise

As the name suggests, the PMTs purpose is to take very small amounts of light (photons), and amplify the signal into something measureable. In this case it is electricity. The PMT consists of an array of metal plates, all with a high voltage between them, such that the electrons are sitting on the edge of their atom, requiring very little excitation to be emitted. When a photon hits the first plate, a few electrons are knocked off, and travels towards the next plate, which then knocks off even more electrons and so on, until it becomes a measureable current. The more signal hits the PMT, the larger the current. Because the PMT needs to pick up on very small amounts of light, it needs to be very sensitive, meaning that very often, the PMT is triggered spontaneously (e.g. by uranium decay). As a result of this, a lot of the measurements are just noise. IceCube then employs a strategy for distinguishing noise from particles, where they look at the surrounding DOMs, in a very specific timeframe, etc.

3.2 The data in IceCube

3.2.1 Simulated data (MuonGun)

To understand the particles in the ice better IceCube simulates muons, neutrinos and noise from well known physical theories and the knowledge about the detector, so it is easier to reconstruct the particles. One simulation of primarily stopped muons is called MuonGun[19] and is the data set which has been used in this thesis.

MuonGun has its name from the idea that it should simulate detector readings from someone shooting a single muon with different starting features into the detector. That is why all "measurements" has features such as "charge_log10" or "time", along with the x, y, and z coordinate of the DOM that would have taken that measurement, as if it was real data. The muongun data set we have used for this thesis consisted of 1.5 million muons with an approximate 70% of the muons being stopped and approximately 30% being through. The energy is distributed around a cluster of the mean energy value of $log_{10}(E) = 2.60GeV$.

The zenith angles are distributed in the $[0, \pi]$ interval with a peak with a mean at 0.667 radians with only a small amount in each tail of the distribution as the muons are harder to detect from directly above (due to the detectors being on the bottom of the DOM) and from directly under the detector (due to the bubble ice columns). One would think that the azimuth distribution should be a uniform distribution in the interval of $[0,2\pi]$ however as the MuonGun data is simulated to look like real data it has six troughs, because of the aforementioned corridors.

3.2.2 Properties of the ice

The Ice on the south pole consists of centuries of compressed layers of ice, but also has impurities. Especially one large layer of dust obscures the ice, hindering the propagation of photons, and making measurements less accurate. Centuries of compression, and the many tons of ice above, has caused any air bubbles in the ice to disappear, but to submerge the DOMs in the ice, it was necessary to melt the columns, and let them re-freeze, creating small bubbles of air around the DOMs, also obscuring the ice. Furthermore just the fact that the detectors have been frozen for so long, mean that it is unknown if they still has the same calibration as they did when they were lowered in the ice. All these issues concerning the ice properties are not well known and is making it harder reconstructing particles in the detector.

4 Introduction To Machine Learning

In this thesis, we have made use of deep learning by making multilayered classification and regression neural networks. The machine learning theory will be covered in this section.

4.1 Machine Learning and Its Benefits

In large complex datasets, there is information that is impossible, or at least cumbersome to extract without the help of complex computer programs. This is the basis of machine learning. By giving the program a set of rules or parameters, and some data will be able to train itself. The idea is that we can analyse data only by building an automated model and from that the model will be able to "teach" itself to recognize different patterns[20].

4.2 Neural Networks

The neural network (NN) takes name after the network of neurons found in any brain, because they both consists of many small neurons, sending signals to eachother.

The neuron takes some vector $\mathbf{X}$, applies weights $\mathbf{w}$ and a bias w_0[3] to it, and finally applies an activation function ϕ, to produce an output $\hat{y}$:

$$\hat{y} = \phi\left(\mathbf{wX}\right) = \phi\left(\sum_{i=0}^{N} w_i x_i\right) \tag{4}$$

When using an activation on the output layer, it decides what kind of values you want. The choice of activation on the output layer, dictates the range and type of the output. When

[3] $x_0 = 1$, so $w_0 x_0 = w_0$

predicting probabilities, one can use an activation like the sigmoid function[4], that only produces outputs in the interval [0,1], just like probabilities.

A simple NN as seen in figure 9 consists of an input layer where the neurons receive the data, then a hidden layer where the neurons process the data and the last layer outputs the final result. All this is happening according to a specific architecture which have been described beforehand[21].

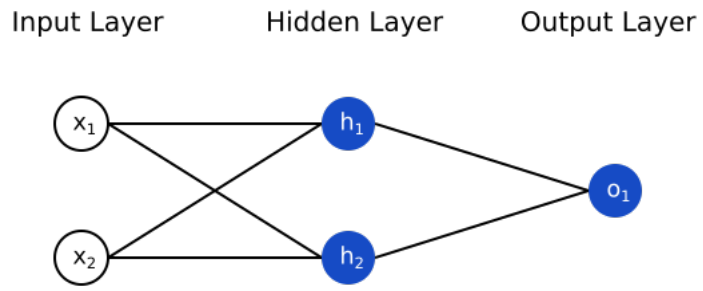


Figure 9: A simple neural network with 3 layers. Image from[22]

4.3 Loss functions & gradient descent

In machine learning, the "learning" part is done via the loss function, $\mathcal{L}(\mathbf{w})$. The purpose of the loss function is to penalize the NN for producing bad outputs[5]. The learning done by the network is actually just trying to minimize the loss, by changing the weights in a way that leads to good outputs.

4.3.1 Loss functions

To explain loss functions, let's look at e.g. the mean squared error (MSE), also used in this thesis for energy regression network:

$$\mathcal{L}(\mathbf{w}) = \frac{1}{N}\sum_{n=1}^{N}(y - \hat{y}(\mathbf{w}))^2. \tag{5}$$

such that for a correct guess $\hat{y} = y$, $\mathcal{L} = 0$, and the model is not punished, and weights are not changed. This loss function goes as x^2, so the severity of the error goes as the error squared, so the function will prioritize correcting large errors. The mean absolute error (MAE) only looks at the absolute value of the difference in prediction versus target, and thus the loss scales linearly, and it is lenient regarding large outliers. Although these two loss functions are very similar, they are not always equally suited for the task at hand. One should always consider what kind of network, and what predictions are desired, before choosing a loss function.

The classification model used in this project applies the Binary Cross Entropy loss (BCE) function,

$$\mathcal{L}(y, \hat{y}(\mathbf{w})) = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i)) \tag{6}$$

where y_i is the i'th label (0 or 1), and $p(y_i)$ is the probability of y_i.

The target values for angle regression lie on a circle. Azimuth is periodic, so $\text{MSE}(0, 2\pi) \neq 0$, and we have to construct a different loss function. We transform truth from a value in the interval $[0, 2\pi]$ to a vector from origo to a point on the unit circle

[4]Like we did in our classifier

[5]i.e. making a bad prediction, wrong classification etc.

$$f(\theta) = \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} \tag{7}$$

Then the computer would understand that $0 = 2\pi$, since $[sin(0), cos(0)] = [sin(2\pi), cos(2\pi)]$. Then we could define our loss, by making the reverse transformation,

$$\mathcal{L}(y, \hat{y}) = 1 - \cos\left(\arctan\left(\frac{\sin y}{\cos y}\right) - \arctan\left(\frac{\sin \hat{y}}{\cos \hat{y}}\right)\right) \tag{8}$$

effectively

$$\mathcal{L}(y, \hat{y}) = 1 - \cos\left(|y - \hat{y}|\right), \tag{9}$$

a periodic function with minima at $|y - \hat{y}| = 2n\pi$, and maxima at $(2n + 1)\pi$ for $n \in \mathbb{N}$.

4.3.2 Gradient descent

When a model has computed the loss, it uses back propagation to find the gradient of the loss function with respect to each parameter. As the goal is to reduce the loss, the network will adjust each parameter in the opposite direction of its gradient. The model will increase and decrease these parameters with step of a size determined by the hyperparameter learning rate. The issue here is that often there will be multiple local minima that it is possible to land in, so the model thinks it has found the best solution, when in fact there possibly exists a much lower global minimum at some other point in the parameter space. A learning rate too small would make it hard for the model to escape local minimas, and a learning rate too high would make it difficult to find the exact point in the region surrounding the global minimum.

4.4 Temporal Convolutional Neural Network (TCN)

Convolutional neural networks (CNNs) are able to detect patterns in space, and 2D CNNs are thus well suited for e.g. image-recognition, and are called convolutional due to its window-sampling way of working, also known as convolving. 2D CNNs will often have many filters, where each filter learns to recognize specific patterns, e.g. horizontal lines, circles, curved lines. Finally, it uses all layers, and if it recognizes two circles separated by a triangle, it may think the image is of a bicycle, since it has seen that combination of shapes in other images of bicycles.

Another type of neural networks is the recurrent neural network (RNN). RNNs are capable of recognizing temporal information in sequential data, as it is able to "remember", in the sense that it can use previous inputs on later inputs. RNNs can also be unidirectional, like the TCN, so that only past elements of the input sequence can affect the current element, and not future.[23]

Where a Recurrent Neural Network (RNN) detects patterns in a time series, and CNNs detect patterns in space, TCNs combine both these features, to detect patterns in time and space.It has been shown that convolutional networks can achieve better performance than RNNs [24].

The RNN processes its input sequence one element at a time, whereas the TCN can process the entire sequence simultaneously, because the same filter is used in every layer, which also cuts down on memory usage. [25]

4.4.1 Dilated convolutions

A characteristic of the TCN is that it uses dilated convolutions. Let k be the size of the kernel[6] and s be the stride[7]. If presented with an array of data, it would then initially look at the k first elements of the array, and then perform some algorithm on those k values[8], to produce a value

[6] i.e. the amount of data points that we look at in each convolution

[7] The size of the steps between each convolution

[8] e.g. the average

fed into a node in the next layer, and then slide s points over to the next k points, and repeat. This convolution would have a dilation of 1, meaning the next data point to consider is the one that is 1 away. Now every node in layer 2 depends on three consecutive nodes in layer 1. So if layer 2 convolves with a dilation of 2, the kernel size remains 3, but it only looks at every other data point, giving it a receptive field[9] $f_r = d(k-1)+1 = 5$.

If using a kernel of size 5 and the commonly used dilation base $b = 2$ such that the dilation $d \in \{1, 2, 4, 8, 16, 32\}$, layer 6 would have a receptive field of $32 \cdot 4 + 1 = 129$, without having any holes.

It has basically the same benefit as increasing the kernel size or filtering, but with fewer parameters, allowing it to be more efficient. You do not want to dilate too aggressively, as that will lead to holes in the receptive field, and some data points will be unaccounted for.

TCNs are causal, meaning that information only travels "forward" in the network, unlike RNNs where e.g. layer n can feed data back into layer $n-1$. Here the i'th element of the output sequence only depends on the 0-ith element in the input sequence.

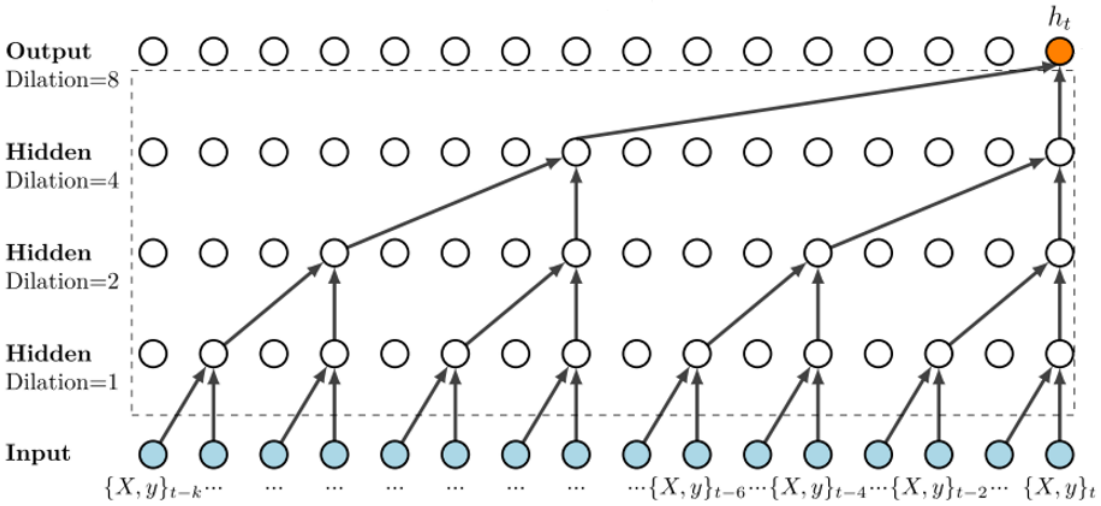


Figure 10: An example of a five-layered TCN with dilation base 2. X is the input data, and y is the input labels. Image from[26]

4.4.2 Padding

TCNs demand that all samples must be the same shape. To include all of the data, one would have to find the event with the most pulses, and then zero-pad all other events, to match that length. For the MuonGun data set the median of the pulses per event was found to be 59, and 99% of events had 232 pulses or fewer, but unfortunately, there were a few large outliers, with the largest event having 5550 pulses. Usually, omitting one percent is no problem, but here, that one percent represent a very distinctive section of the data. Due to limited processing power, we settled on only taking events with 250 and fewer pulses, corresponding to 99.3% of the data. For events with more than 250 pulses, we sampled 250 samples randomly, in hopes of not losing any ability to predict large events. If we wanted to gain that last percent data, one would have to process 24 times the data. Due to how the TCN works, it is not possible to make predictions on events larger than those it has been trained on, so we can not test the effects of our pulse-selection this way.

[9] Equation valid for $k > 1$

5 The process of creating neural networks with different purposes

A large part of this project was centered around creating a working NN in the shape of a TCN and be able to make classification and regression on stopped muons in the IceCube simulation. Therefore it will be discussed how the process went and how the model and all the work before the model would be able to run has been made.

Two of the models in this thesis was centered around the zenith and azimuth angles which tells the angle of entry of the muons, allowing the model to point where in the sky the muon came from, but the actual purpose of making these models is that we in the future can compare it to real data to see how well we are able to simulate the properties of recorded events in the detector. In IceCube the stopped muon is the event we are able to model the most precisely, due to sheer volume of data. With this as the basis we are more capable of simulating data, and comparing it to real data, allowing to possibly modelling of neutrinos to be more precisely in the future. As IceCube detects many thousands more muons than neutrinos. Meaning that without (stopped) muons, we would have a much thinner foundation for building ML models to predict neutrinos.

The product of the thesis is one classification network and three regression networks:

Binary classification: Stopped and through muons (SMC)

Regression: Muon energy (Ereg)

Regression: Muon zenith angle (Zreg)

Regression: Muon azimuth angle (Areg)

The three regression models were trained on stopped events, as they were meant to be used on stopped events.

It needs to be mentioned that all the models in this thesis have been constructed using keras-tensorflow where the model foundation comes from keras-TCN[27] which we have created the TCN-layers from.

5.1 Purposes for the models

For the classification of stopped muons, the purpose was to get a model which would be able to classify whether or not a muons stopped or went trough the IceCube detector. Meanwhile the purposes for the three different regression models was to be able to predict the energy, the zenith angle and the azimuth angle all for stopped muons. All the models had to be set up with different parameters and some even needed to have the input data modified in order to get better predictions which all will be covered in the following.

5.2 Architecture and general parameters

5.2.1 Dataloader

In the process of creating the neural networks we quickly learned that the computers used in this thesis did not have the necessary memory to process data sets as large as MuonGun and it was necessary to use a dataloader. Meaning that we could break the data into smaller batches, and feed them to the model one batch at a time, which is what we call a dataloader. In this thesis case we found it most useful to use a dataloader which used a batch size of 1.000 meaning that the model received 1.000 different events at a time N times where N was the total number of events divided with the batch size.

5.2.2 Events in the batches

Muongun which was mentioned as the data set used in this thesis has approximately 70% stopped muons, and approximately 30% through in it with a total of 1.5 milions simulated muons. To

ensure a balanced data set for the SMC we first tried to only take as many stopped muons as we took stopped, i.e. a 50/50 split. This constraint meant we could only look at approximately 60% of the data set, which made it not worth it performance-wise as we wished to look at even more muons, so we dropped this idea. Instead it was found that it could be better to shuffle the events, so they were not given locally for the detector in each batch, meaning that we would have events from all over the detector in one batch instead of just for the same area. This of course still gave an approximately 70/30 distribution of the stopped and not stopped muons but it was quickly realised that the model were able to learn which muons stopped and which did not with this distribution. The three regression models only adapted the shuffling of the events as they were only focusing on the muons which are stopped in the detector.

5.2.3 Features and truths for the models

The truths for each of the models has already been discussed to some degree as they are pretty self explanatory meaning that the SMC of course looks at stopped muons (stopped_muon), the Ereg looks at the energy (energy_log10) and the angle regressions (Zreg and Areg) looks at the different angles (zenith and azimuth) in the truth values.

However the features for all the models was a little different process. It has to be mentioned that all the models had the same features as the features was found usefull for all the models. First they were chosen to be the location of the muons, time and the charge recorded in the pulses meaning an array of (dom_x, dom_y, dom_z, time, charge_log10) (table 2). These actually worked decently good but later in the process after working with the different models and getting to understand them better it was found useful to implement two new features which were pulse_width and SRTInIcePulses. These two features involved a lot of "hidden" information of the muons which the model was able to read and understand. This "hidden" information was for pulse_width a variable which were either 1 or 8 depending on whether the DOM took a recording every 1 or 8 nanosecond. And the feature SRTInIcePulses might be even more important which tells if a measurement in the event would have survived an SRT cleaning or not meaning that the models were told if IceCube would have taken the measurement as noise or not. Which is important in the aspect that the model should be able to understand wheter or not a measurement could be noise or not.

5.3 Stopped muon classification

In the early stages, the classifier would only predict one constant decimal number for all events, instead of correctly predicting the label 0 or 1. We solved the problem by letting the label 0 be represented as [1,0], and the label 1 represented as [0,1] (one-hot encoding). This way, the model would output two numbers, [the probability of 0, the probability of 1].

The threshold of classification was 0.5, meaning the model had to predict a probability of over 0.5 on stopped, to be classified as such. The classifier then makes a prediction of $[p(y = 0), p(y = 1)]$, where $p(y = 0) + p(y = 1) \approx 1$, so $1 - p(y = 0) \approx p(y = 1)$. It does not exactly add to 1, due to rounding.

Before the final model was constructed multiple classifications of stopped muons were done and in the "early" process of the project when the TCN-model started working we tried to predict stopped muons which was not as good as one would have liked. And therefore all the different parameter changes and adjustments for the model as described above was done.

The loss function for the stopped muon classifier was chossen to be BCE described with equation 6 with different activation functions in the layers as seen in table 1 together with the the other parameters.

The model itself ended up being trained on 1 milion events distributed as mentioned in section 5.2.2 and the results for the model can be seen in section 6.1.

5.4 Muon energy regression

For the energy regression, we used MSE as the loss function as described with equation 5.

The energy target values were in an interval spanning from negative values to larger than 1 values. Therefore we needed a final activation function similar to the ReLU, that allowed for values above 1, but also for negative values. However an activation function that is linear for both $x < 0$ and $x \geq 0$, is effectively the same as not applying an activation function, so that's what we did. However in all the other layers there was used the elu activation function.

The model ended up being trained on 500.000 events where the muons was stopped as this model was only centered around stopped muons as mentioned earlier.

5.5 Muon angle regression (zenith and azimuth)

As mentioned in section 4.3.1 about loss functions we have used a "homemade" loss function.

This model was a little bit easier to get rolling as all the hard work on making the TCN work was already done when classifyning stopped muons and predicting energy. And with this setup the zenith prediction as shown in section 6.3 was made and the azimuth prediction shown in section 6.4 was made.

The final activation layer for these two models was chosen to be tanh after some optimization showed us that this was the best choice for the combinations we tried.

The events was chosen the same way as they where for the energy regression as the angles also only was centered around stopped muons for this thesis meaning that the model trained on 500.000 stopped muons.

Table 1: The different parameters used for the final models in the thesis.

Parameters	SMC	Ereg	Zreg	Areg
Loss	BCE	MSE	sincosloss	sincosloss
Start activation	LeakyRelu	elu	LeakyRelu	LeakyRelu
Mid activation	Relu	elu	LeakyRelu	LeakyRelu
Final activation	Sigmoid	elu	Tanh	Tanh
Layers	4	4	4	4
Epochs	88	58	191	90
Number of filters	16	32	32	32
Kernel size	10	3	5	3
Output dimensions	2	1	2	2
Learning rate	1e-3	1e-3	1e-3	1e-3
Events trained on	1,000,000	500,000	500,000	500,000

5.6 Evaluation of models

5.6.1 Classification

In binary classification the receiver operating characteristic (ROC) curve is a measure of how well the correct and wrong predictions are separated (see figure 11). The curve is calculated by integrating from 0 up to 1, and all the correctly classified points go towards the true positive rate (TPR), and the wrong go towards the false positive rate (FPR). Ideally, they would be separated completely, and FPR would stay at 0 until the FPR is 1. The closer to the corner the curve gets, the better, and thus the area under the curve (AUC) also goes towards 1, the better the separation is. For simplicity, the (AUC) is used to evaluate the curve.

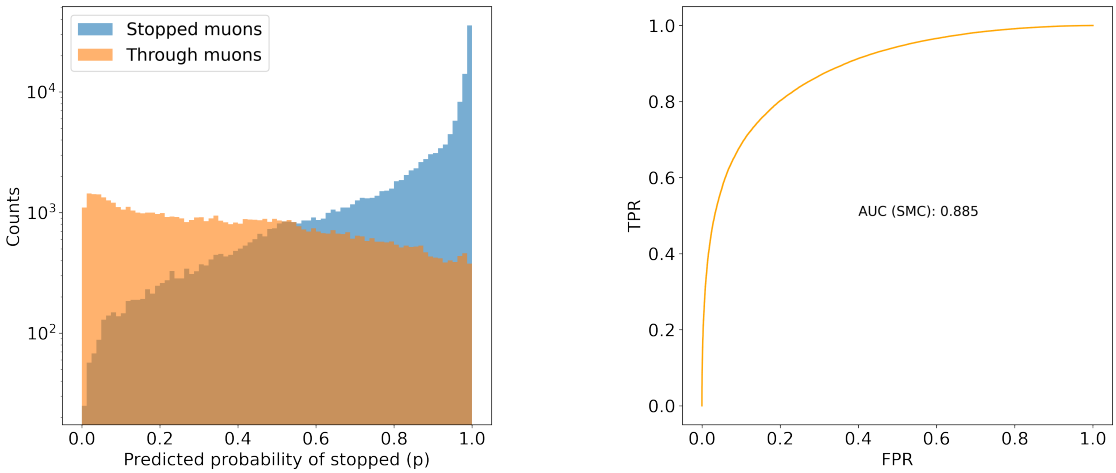
5.6.2 Regression

For regression, we can not say that the model was right or not, only how close. The metric we used to measure this was that we plotted a 2-dimensional histogram of the target values and the predicted values. As these two values ideally would be the same, they would form a diagonal line like $f(x) = x$, and thus have a correlation of 1, such that a change of α in x, would correspond to a change of α in y. A correlation of 0, means that there is no dependency between x and y. Therefore correlation proved to be a neat scale of 0 to 1, showing how well our regression models predicted.

6 Results

In this section we will present and discuss various plots illustrating the produced predictions of each of the four network separately, compared to the target values that the simulated data were based on.

6.1 Stopped muon classifier



(a) (b)

Figure 11: (a) Histograms of the probabilities of being stopped for mouns that are actually stopped (blue), and muons that are actually through (orange). (b) ROC curve showing the true positive rate (TPR) versus the false positive rate (FPR) for the stopped muon classifier.

The performance of the SMC (figure 11 (b)) shows an AUC of 0.885 with the distribution of the prediction scores shown as in the histogram figure 11 (a). This tells us that the model is capable of predicting whether or not a muon i stopped in the detector.

However we can also see from figure 11 (a) that the model is fairly confident in its predictions, with a large peak for the blue histogram at $p(1) = 1$. It shows that the model in general is most certain when its predicting stopped for muons which actually are stopped according to truth. Where it is not as certain for the muons that are actually through (figure 11 (a)). We can tell from figure 11 (a) and (b) that the model is capable of predicting stopped muons but there is still room for improvement as the AUC can go even higher.

First of all it is not necessarily the perfect parameters which have been used in this classification, it was simply the one we found best due to trying multiple different parameter combinations (the parameters used can be seen in table 1. As mentioned in section 4.4.2, events with more

pulses than 250 were not fully represented, and we therefore do not know how well the model handles them.

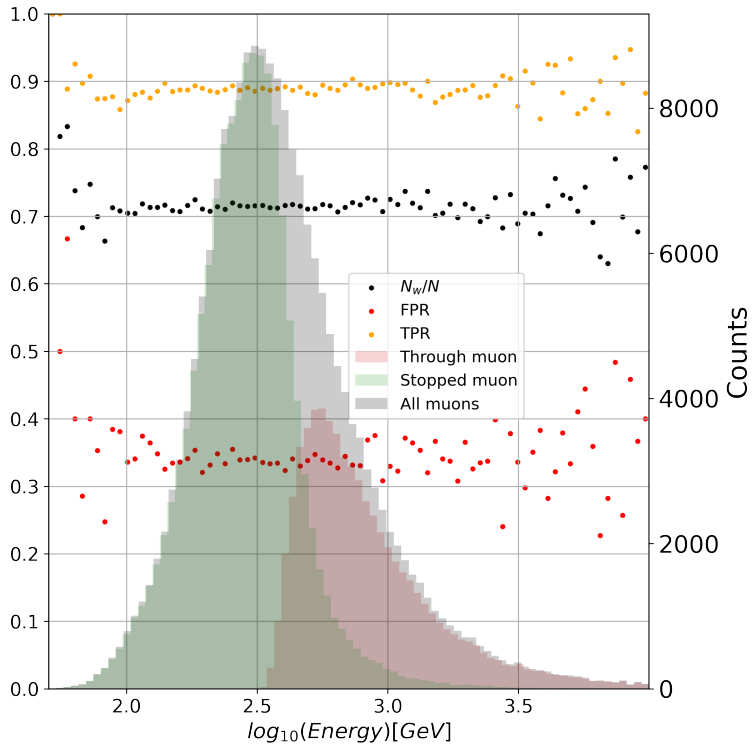


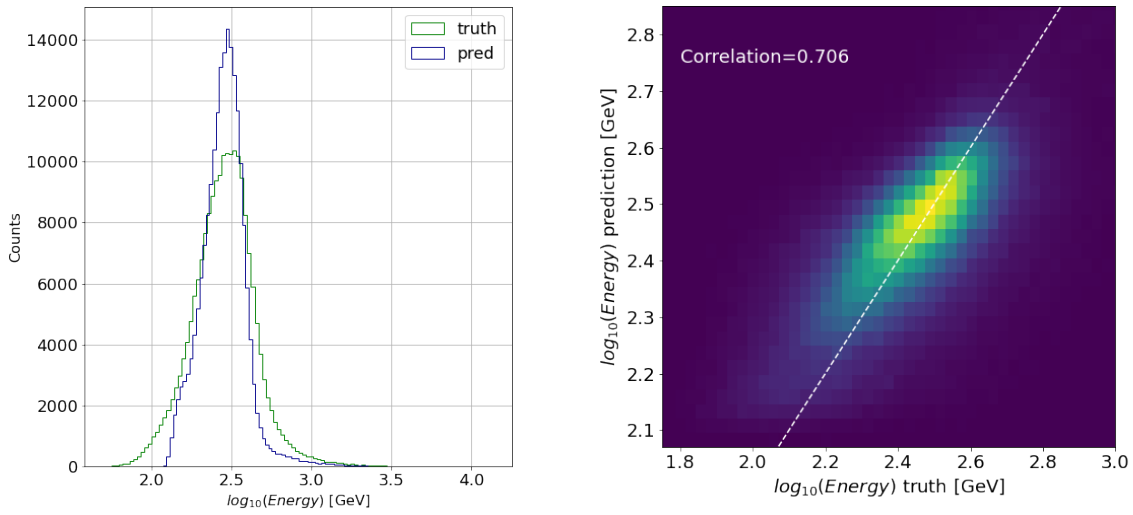
Figure 12: In grey: A histogram of the energy of all the events that has been predicted on. In green and red, histograms of the actually stopped and through events respectively. All points correspond to the bin they share x-coordinates with. The black points are the ratio of correctly classified events to the total number of events in the bin. The orange points are the true positive rate, and the red points are the false positive rate.

Figure 12 shows that the SMC in general is consistent in predicting stopped and through muons as a function of the energy. The TPR is fairly accurate (almost a straight line through 0.9) which tells that the model is good at predicting correctly when the muons is actually stopped. The FPR is also fairly decent but not as good as the TPR as one would expect it to be closer to 0.0 if the model was really good at predicting through muons as through. However we would have expected the energy to have some sort of influence on how good the model would be at predicting stopped muons as the distributions of stopped and through are separated fairly well. If one had a perfect energy estimator, one could make a decent classifier by simply classifying all events with energy less than $10^{2.5}$ GeV. So we postulate that the classifier could be improved greatly if it was possible to put the Ereg model inside the classifier.

The black points show the ratio of correctly classified events in each bin. Like the TPR and FPR, they do not appear to depend greatly on the energy of the events. The points of the bins with low counts fluctuate a lot, as one could expect with such low statistics. However overall this still shows that the model has room for improvement just like 11 (a) and (b) e.g. with the improvements mentioned earlier.

6.2 Energy regression

The Ereg has been validated according to the truth values to see how close the model was in its predictions compared to the truth values. This section will then present and discuss the results of the energy regression model and its predictions.



(a) (b)

Figure 13: (a) Histograms of the actual energy of the muons and the predicted energy of the muons. (b) A 2D histogram of the actual energy of the muons and the predicted energy of the muons.

The Ereg predictions has a correlation of 0.706 and the distribution of $log_{10}(truth) - log_{10}(prediction)$ has a $\sigma = 0.144$ GeV when comparing the predicted energies to the truth values. Figure 13 (a) also shows that the predicted energies follows the same shape as the truth values of the energies but the predicted values does not reach the extremas of the truth values, but prefers to predict values close to the peak of the actual distribution (E=$\approx \log_{10}(2.5)$ GeV). The reason behind the model being worse at predicting the extremas could possibly be in the loss function. As mentioned earlier in section 4.3.1 the loss function "punishes" the model with higher losses when it is predicting far from the truth values and when most of the data is centered around E $\approx \log_{10}(2.5)$ GeV, the model will on average get the smallest losses by just predicting close to that area. When it is predicting on the extremas it has a minimum of truth values to learn from and therefore it will automatically find it harder to predict those values as it has a minimal amount of knowledge of the data in the extremas. A way to try and make up for this would be to give the model data which were evenly distributed with no significant peaks as the model then would have a lower chance of getting higher losses in certain areas of the distribution.

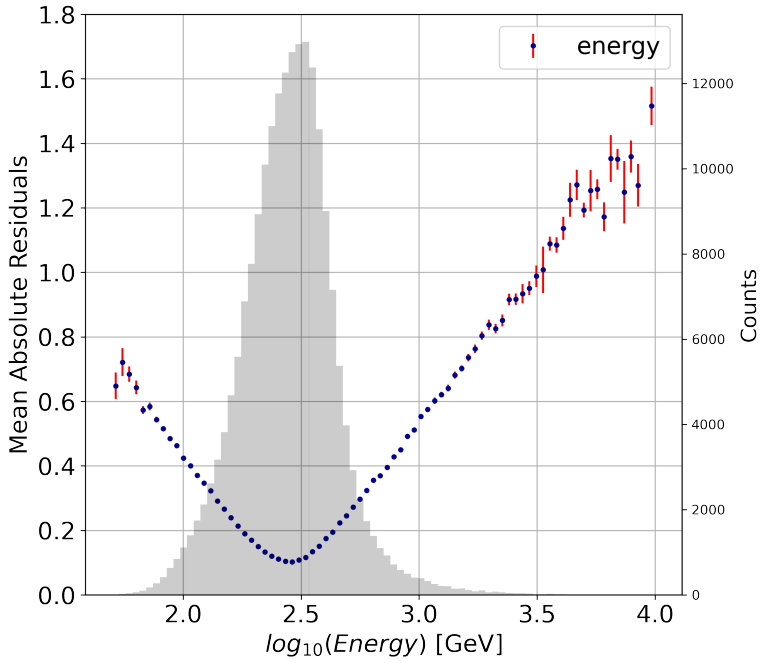
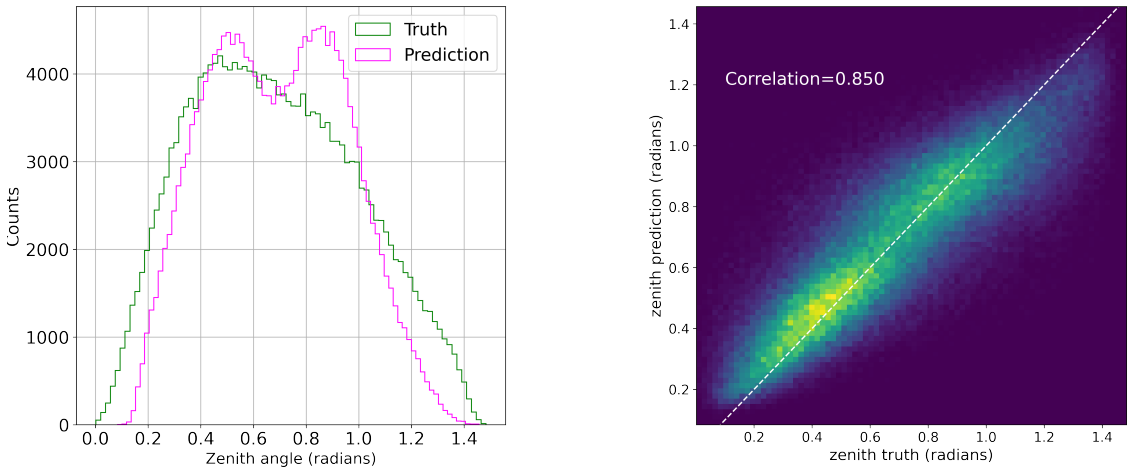


Figure 14: Mean absolute residual of each bin in the histogram of the distribution of the energy for the stopped muons. The error is the standard deviation of the bin divided by $\sqrt{N}$.

In figure 14 it is shown that the mean absolute residual has a minimum close to $log_{10}(E) \approx 2.5$ GeV. That is probably because that when the model is uncertain, it guesses $log_{10}(E) \approx 2.5$ GeV and because most events have an energy close to that, on average, it will not be far off. Also as mentioned before when discussing figure 13 the model does not like to guess near the extremas which may be due to the losses from the loss function which already has been explained. However this might not be the only problem as the same counts for this model as it did for the SMC that the parameters chosen might not be the best yet and needs to be optimized further. However the results still shows that the model is predicting the energies especially in the area where the energy is distributed around.

6.3 Zenith regression

The Zreg model has been validated according to the truth values to see how close the zenith regression model is in its predictions compared to the actual truth values. This section will then present and discuss the Zreg model and its predictions.

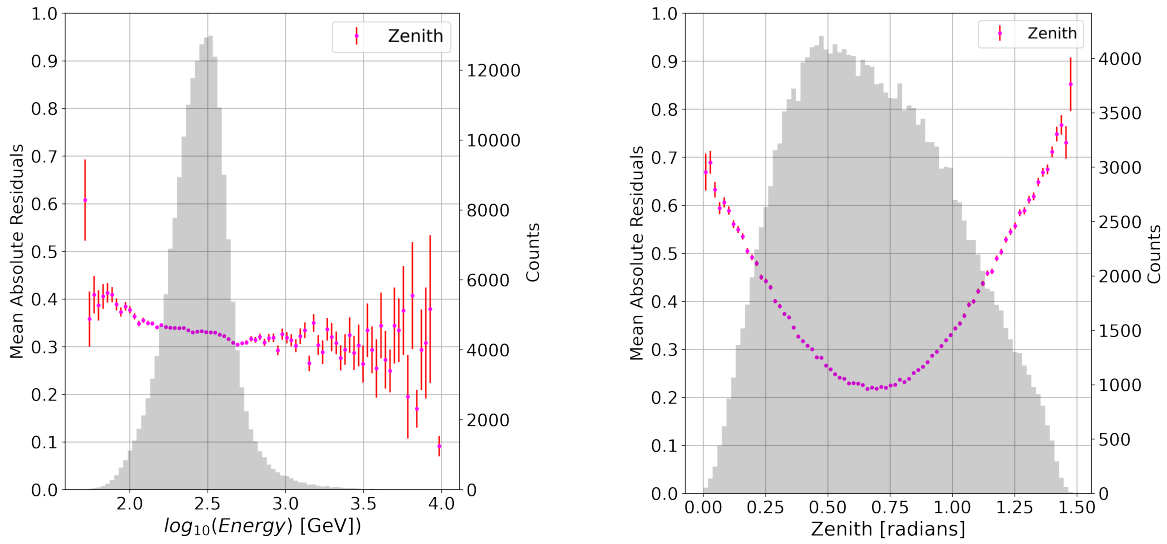


(a) (b)

Figure 15: (a) Two histogram of the actual zenith angles and the predicted zenith angles, respectively. (b) 2D histogram of the actual zenith angles against the predicted zenith angles. The correlation between truths and predictions is denoted in the top left corner of (b).

As seen from figure 15 (a) the distribution of the predicted zenith angles of the zenith regression model looks somewhat like the distribution of the true values, except that it has two peaks. This model uses the sincosloss function, that predicts $sin(\hat{y}), cos(\hat{y})$ rather than $\hat{y}$, and it is suspicious that the predicted distribution has two peaks close to $tan(0.5) \approx 0.54$ and $cos(0.5) \approx 0.88$.

Even though the distributions look alike, it does not tell us exactly how precise the individual zenith angles are predicted but it tells us that the overall distribution is decent. From figure 15(a) we can also see that the model in general is bad at predicting the extrema, and makes almost no predictions in the intervals [0,0.1] and [1.4, $\pi/2$]. Figure 15 (b) shows that the predicted values is fairly close to the truth values as the 2D histogram follows a linear pattern. We found the correlation of the predicted zenith angles with the zenith regression model to be 0.850, making it the model, that has the highest correlation between prediction and truth of the four. The sigma of the distribution of the difference in truth and prediction was calculated to be $\sigma = 0.166$ radians. The model is decent, but could possibly be better with optimizing the parameters. A lot of the same parameters were used from the stopped muon model, and perhaps there exists a set of parameters that would be great for this model, that was overlooked, because it was poor for the classifier.



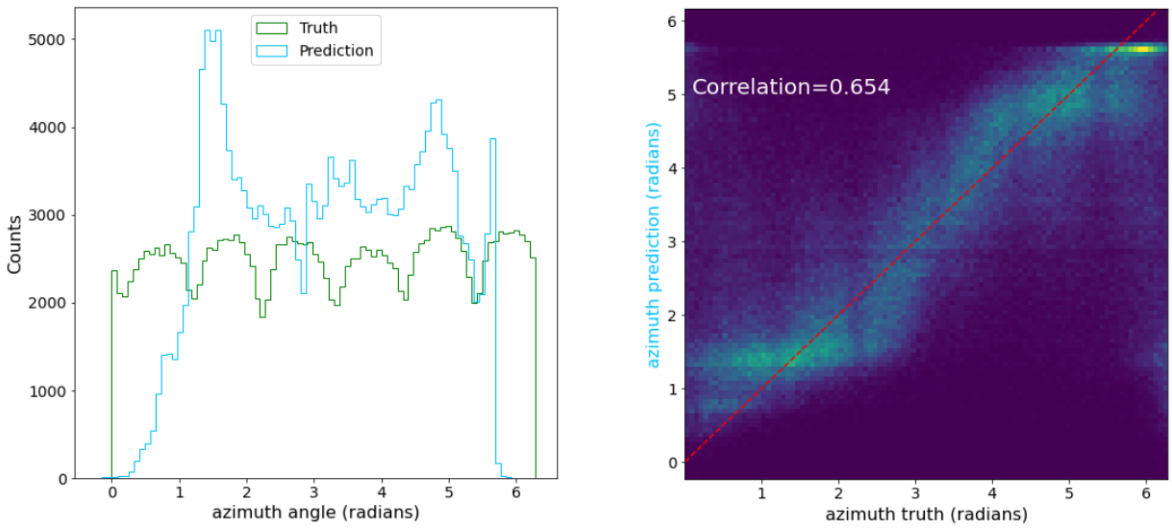
(a) (b)

Figure 16: (a) shows mean absolute residual in each bin of the histogram of the distribution of the energy for the muons. Figure (b) shows mean absolute residual in each bin of the histogram of zenith angles of the muons.

Both figure 16 (a) and (b) states the obvious, that the error is lowest in the bins with the highest counts. However figure 16 (b) also shows that the lowest mean residual is found to be 0.22 at the center bin with a zenith angle value of 0.659 ± 0.015 radians ($37.76° \pm 0.86°$). Again, the model has the lowest mean residual at the value of the peak of the energy histogram.

6.4 Azimuth regression

The azimuth regression model has been validated the same way as the zenith model and its rsults will be discussed and presented in this section.



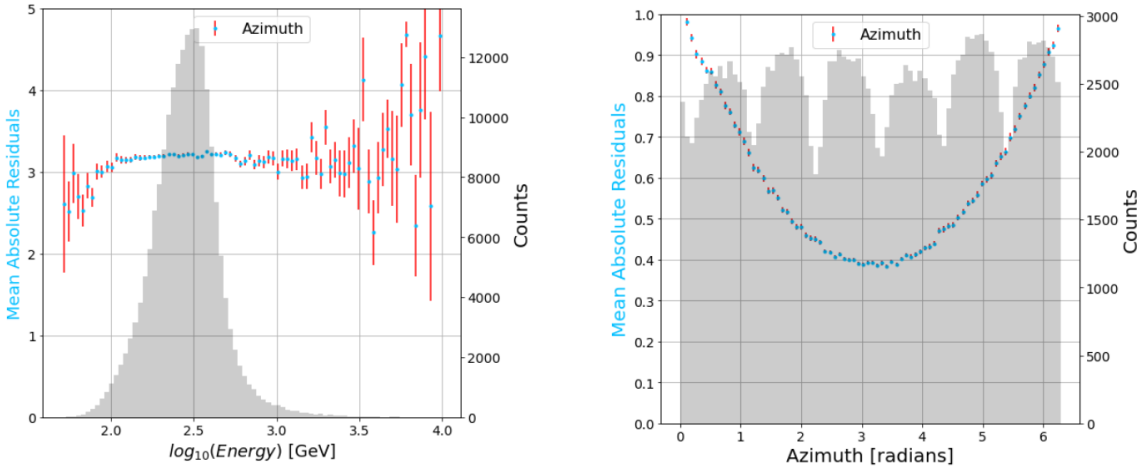
(a) (b)

Figure 17: (a) Histograms of the actual azimuth angle of the muons and the predicted azimuth angle of the muons. (b) Histograms of the actual azimuth angle of the muons and the predicted azimuth angle of the muons.

It is quickly observed from figure 17 that the Areg is not that good at predicting the azimuth angles. The correlation between the predicted angles and the truth values for the angles is 0.654 which is worse than the two other regression models. The sigma of the distribution of the difference in truth and prediction is 1.396 radians, which is very high, but it is also due to azimuth being a periodic value as previously discussed, and is a rather poor metric here. Once again it is observed that the model is reluctant to predict values at the extremas. This could be because there is a potential for higher losses at the extremas, so the model takes the safe way out, and predicts a value closer to the middle, the same thing as mentioned for the energy regression model.

As previously mentioned, the output of the model was a vector $[\sin(\hat{y}), \cos(\hat{y})]$, and to extract $\hat{y}$ from the output, we used the function $prediction = \left|arctan\frac{sin(output1)}{cos(output2)}\right|$. This function worked well for zenith, because it is linear such that $f(x) = x$ in the interval $[0, \pi/2]$, but not for values above $\pi/2$.

In figure 17(b) the dense areas in green looks like some sort of trigonometric function, hinting at the fact that it is the trigonometric loss function, that is not perfectly suited for this task. There are also 6 (maybe 5. The leftmost is very faint) visible vertical black lines, that corresponds to the 6 troughs in figure 17(a).



(a) (b)

Figure 18: (a) shows a histogram of the distribution of energy in the dataset. For every bin in the histogram, the mean absolute difference of prediction and target value is plotted as a dot. The error on the dot is calculated as $\sigma/\sqrt{N}$. The large errors to the left and right of the peak are due to low statistics. (b) shows the same as (a), but the histogram in question is of the distribution of azimuthal angles.

Figure 18 (a) and (b) both show the same obvious as figure 16 did for the zenith angles that the error is lowest in the bins with the highest count.

The model especially had difficulties with predicting values at the extrema, which is why the mean difference in truth and prediction are so large, that the prediction is as good as random. This model is very inaccurate, and would need some modification, in order to be good. As the biggest difficulty is posed by the periodic nature of the distribution, one should probably look to the loss function, in order to improve it.

6.5 Speed

The current reconstruction algorithm that IceCube uses (**RetroReco**) takes 5-40 seconds to make a reconstruction[1]. Our models take between $1.5 \cdot 10^{-4}$ and $1.5 \cdot 10^{-4}$ seconds to predict one value, using an NVidia GeForce RTX 2070 GPU. We could not obtain any reconstructions

made by RetroReco on the muongun data, so we can not compare accuracy, but we can tell that the TCN models are a lot faster than RetroReco if our prediction time holds for other data sets too.

7 Discussion and Conclusion

In this thesis we have tried to predict simulated MuonGun values from the IceCube observatory with 1 classification model and 3 regression models. We have tried to classify whether the different muons stopped or went through the detector, energy values for the stopped muons and the zenith and azimuth angles for the stopped muons. In the following we will go through what we have learned from the models.

First for the SMC we managed to predict the stopped muons with an AUC of 0.885 (fig. 11) meaning that we can see that the model is able to tell correctly whether the muons are stopped or not. But there is still room for improvements which have already been discussed. We hypothesized that since there is a great tendency for stopped muons to be below $\approx 316GeV$, and through muons to be above, the stopped muon classifier would be best at classifying for low and high energies. However, we found that the classifiers accuracy does not appear to depend on the energies of the events.

The Ereg model predicted the energies for stopped muons compared to the truth values with a correlation of 0.706 and $\sigma = 0.144$ GeV. It was also found that the model in general followed a pattern around $E = log_{10} \approx 2.5$ GeV due to the cluster of data in that area. It was also found that the model found it hard to predict the extremas of the distribution of the energy where the possible reasons behind and possible optimization of this has already been discussed.

For the Zreg and Areg models it was found that they were much better at predicting the zenith angles rather than the azimuth angles. The Zreg model predicted with a correlation of 0.850 and $\sigma = 0.166$ meanwhile Areg model predicted with a much lower correlation of 0.654 and much higher sigma $\sigma = 1.396$. The fact that the azimuthal target value was periodic proved a large stumbling block in trying to predict it accurately. Fortunately the angles is not the most important part of the IceCube detector as energy and stopped muons is way more important in reconstructing stopped muons. But still the Zreg is of importance as it concludes that the detector is able to say that muons are coming from above meaning that they are coming from cosmos.

Overall it can be concluded that 3 out of 4 of the models are capable of doing what we wanted them to do as SMC is predicting stopped muons, Ereg is capable of predicting energies and Zreg is capable of predicting the zenith angles. The Areg is not the best at predicting azimuth angles and need more optimization. In general it can also be concluded that all the models has room for improvement where possible optimizations has already been discussed.

8 Further Research

On a small scale it was seen throughout the result and discussion section that there was room for improvement in all the models which is something we would like to look into in the nearest future. There could be looked further into the parameter optimization and in the case of the zenith regression model there could be looked for a better loss function. As mentioned in section 6.1, we would like to combine the SMC and Ereg models, to produce a better classifier. Another way to test it, would be to have the Ereg's estimate of the energy of an event be a feature in the data.

It could be interesting to apply the models from this thesis to real data, and compare them to IceCube's algorithm RetroReco. In relation to this it would be interesting to have the models (when they have been optimized even further) to look at neutrinos in IceCube instead of muons as they are some the least explored pieces of the standard model. They are really hard to detect

and therefore it could be interesting to see how well the TCN from this thesis would be to reconstruct them.

It would also be interesting to look further into which effect the energy has on the stopped muon classifier as one would expect it to be better at low energy muons but as seen in this thesis it was not the case for this classifier. It would be interesting to see if this was just a coincidence or if TCNs might be able to understand stopped muons better in relation to their energies than first expected.

If we are staying within the scope of this thesis, we can look at one of the reasons that we have looked at stopped muons. We would like to be able to identify stopped muons when they occur, and then have a look at what is measured directly after a stopped muon decays in the ice. This is because muons decay to electrons, electron antineutrino, and muon neutrinos. So if the models were improved, it would in theory be possible to determine the energy of the decayed muons, and thus infer the energy of spawned neutrinos and anti-neutrinos. It would make sense that given the extra information about the energy of the muon from which the neutrino came, one could more accurately determine the energy of the neutrino.

Outside the scope of this thesis, there are some behaviours on the neutrino oscillation that hints at a fourth type of neutrino, the sterile neutrino. This particle would need to be unable to interact weakly (otherwise we would have seen them in neutrino oscillation), meaning they can only interact through gravity. This would make them incredibly hard to detect, as it relies on the weak or electromagnetic interactions.

9 References

References

[1] Rasmus F. Ørsøe, 2021, *A Graph Neural Network Approach to Low Energy Event Reconstruction in IceCube Neutrino bservatory*
Niel Bohr Institute

[2] Peter H. Hansen, *Standardmodellen.*
https://fysikleksikon.nbi.ku.dk/s/standardmodellen/
(visited on 05/26/2021)

[3] Office of science, *DOE explains... the Standard Model of Particle Physics*
https://www.energy.gov/science/doe-explainsthe-standard-model-particle-physics
(visited on 05/26/2021)

[4] Wikimedia Commons contributors, *Standard Model of ElementaryParticles*
https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg
(visited on 05/26/2021)

[5] Hadiseh Alaeian, *An Introduction to Cherenkov Radiation.*
http://large.stanford.edu/courses/2014/ph241/alaeian2/
Visited on 06/12/2021

[6] P.B. Price, K. Woschnagg, *South Pole ice characteristicse.*
https://www.sciencedirect.com/science/article/pii/S0927650500001420?via%3Dihub
(visited on 06/15/2021)

[7] Marquis R. Nave, *The lifetime of leptons*
http://hyperphysics.phy-astr.gsu.edu/hbase/Particles/lepton.html
(visited on 06/16/2021)

[8] The HAWC Collaboration, *Cosmic rays*
https://www.hawc-observatory.org/science/cosmicrays.php
(visited on 06/16/2021)

[9] IceCube-gen2, *The World's largest neutrino detector*
https://www.icecube-gen2.de/project/index_eng.html
(Visited on 06/10/2021)

[10] Denver Dang, *Stopping power diagram*
https://physics.stackexchange.com/questions/156755/stopping-power-diagram
(visited on (16/06-2021)

[11] Blake Stacey, *Supernovas: Making Astronomical History*
https://snews.bnl.gov/popsci/neutrino.html
(visited on 06/16/2021)

[12] B. R. Martin and G. Shaw, *Nuclear and particle physics: an introduction*
(Wiley, 2006) pages 71-79.

[13] Benjamin Eberhardt, *IceCube Overview*
https://icecube.wisc.edu/about-us/overview/
(visited on 04/25/2021)

[14] IceCube Collaboration. *The IceCube neutrino observatory.* https://icecube.wisc.edu/gallery/detector/modulagallery-7032-2064 (visited on 06/01/2021)(cited on)

[15] IceCube Collaboration, *IceCube Detector* https://icecube.wisc.edu/science/icecube/ (visited on 04/25/2021)

[16] R. Abbasi et al. *"The design and performance of IceCube DeepCore", Astroparticle physics* **35**, pages 615-624

[17] University of Wisconsin-Madison Physical Sciences Lab, *Digital Optical Modules* http://www.psl.wisc.edu/projects/large/icecube/more-icecube/dom (visited on 06/16/2021)

[18] IceCube collaboration, *Digital Optical Module* https://icecube.wisc.edu/gallery/diagrams/modulagallery-7055-1605 (visited on 06/16/2021)

[19] Jakob van Santen, *Neutrino Interactions in IceCube above 1 TeV: Constraints on Atmospheric Charmed-Meson Production and Investigation of the Astrophysical Neutrino* Univeristy of Wisconsion–Madison

[20] Unknown author, *Machine Learning What it is and why it matters* https://www.sas.com/en_us/insights/analytics/machine-learning.html (visited on 05/26/2021)

[21] James Chen, *Neural Network* https://www.investopedia.com/terms/n/neuralnetwork.asp (visited on 05/27/2021)

[22] Victor Zhoy, *Machine Learning for beginners: An introduction to Neural networks* https://towardsdatascience.com/machine-learning-for-beginners-an-introduction-to-neural-networks-d49f22d238f9 (visited on 05/27/2021)

[23] IBM Cloud Education, *Recurrent Neural Networks* https://www.ibm.com/cloud/learn/recurrent-neural-networks (visited on 06/15/2021)

[24] Francesco Lässig, *Temporal Cnvolutional Networks and Forecasting* https://medium.com/unit8-machine-learning-publication/temporal-convolutional-networks-and-forecasting-5ce1b6e97ce4 (visited on 02/18/2021)

[25] Raushan Roy, *Temporal Convolutional Networks* https://medium.com/@raushan2807/temporal-convolutional-networks-bfea16e6d7d2 (visited on 06/15/2021)

[26] Barak Or, *Temporal Concolutional Networks, The Next Revolution for Time-Series* https://towardsdatascience.com/temporal-convolutional-networks-the-next-revolution-for-time-series-8990af826567 (visited on 18/02/2021)

[27] Philippe Remy, *Keras-tcn* https://github.com/philipperemy/keras-tcn (visited on 05/30/2021)

[28] Aartsen et al, *Characterization of the Atmospheric Muon Flux in IceCube*
`https://arxiv.org/abs/1506.07981`
arXiv:1506.07981v2

10 Appendix

10.1 How long it would take to produce a large sample of stopped muons with a TCN?

Our classifier has an accuracy of 81%, so in order to produce a dataset of 1 million events with 95% stopped muons, it would need to predict on a data set with $10^6/0.81 = 1234568$ stopped muons. Icecube detects 3000 muons per second[28], if x is the ratio of the detected muons that stop in the detector, that is $3000 \cdot x$ stopped muons per second.

$$\frac{12345679}{3000 \cdot x\ s^{-}1} = 4115 \cdot x^{-1}\ s$$

if we say that 200k predictions of one feature takes 35 seconds, and for one reconstruction, we need to predict 4 values, then we would need to add

$$35 \cdot 4 \cdot 6.17 = 864s$$

to the total time.

In total, the time need to detect and process enough data to produce a sample of 1 million muons with 95% stopped muons, can be described as a function of the ratio of stopped to through, called x, as such:

$$t = 4115x^{-1}s + 864s$$

10.2 Specifics on IceCube data

Table 2: Features in MuonGun with description and datatype where each entry represent an activated PMT.

Name	Description	Type
Event no.	The designated event number	
string	The string number of the activated PMT	int
dom	The DOM number of the activated PMT	int
pmt	The PMT number of the activated PMT	int
dom_x	The x location of the activated PMT	float
dom_y	The y location of the activated PMT	float
dom_z	The z location of the activated PMT	float
pmt_x	The x component of the PMT	float
pmt_y	The y component of the PMT	float
pmt_z	The z component of the PMT	float
pmt_type	Type of PMT (IceCube, mDOM, D-Egg)	int
time	The relative time coordinat of the hit	int
charge_log10	The charge recorded in the pulses in $\log_{10}$	float
pulse_width	Pulse width, (DAC-dependent)	int
SplitInIcePulses	Pulse included in SplitInIcePulses cleaning	bool
SRTInIcePulses	Pulse included in SRTInIcePulses cleaning	bool

Table 3: Truths in MuonGun with description and type where each entry represents a monte carlo generated particles.

Name	Description	Type
Event no.	The designated event number	
energy_log10	Energy of the particle in log_10	float
time	The relative interaction time of the particle	float
position_x	The x-component of the interaction position	float
position_y	The y-component of the interaction position	float
position_z	The z-component of the interaction position	float
direction_x	The x-component of particle pointing vector	float
direction_y	The y-component of particle pointing vector	float
direction_z	The z-component of particle pointing vector	float
azimuth	Azimuthal direction of particle	float
zenith	Polar direction of particle	float
pid	Particle ID	int
interaction_type	Type of interaction (neutral or charged)	int

10.3 Testing a model

In machine learning and coding in general you will once in a while and for some mostly have to work with a large data set. For the ones who has tried to work with a large data set once or twice can tell that there is a possibility that it takes a significantly time to run larger amount of data trough a model. The larger and more complicated the data set is the more time the model needs to work through it all. When you have something that takes a long time to run through its almost every time a good idea to run a small test experiment as most physicist/scientists learns through their studies. For example lets say we want to send a rocket to space. First of all it is very expensive but also very time-consuming to design and make calculations for this project. Therefore it is a good idea to design a smaller experiment where you just want to make sure that the rocket can reach a certain height without problems. The same goes for machine learning. Before we wanna run thousands of events with millions of numbers through a model we wanna now if the model will be able to recognize a simple experiment. A simple experiment could just be 100 events where we give each event one or two features which is a number between 0.00 and 0.99. The simple experiment of course needs to have similar traits to the real experiment therefore we could give each event a truth value of either 1 or 0 for whether it is stopped or not if we want to simulate stopped muons. To test the model we could then train the model on all the events where we have the two features and the truth value. After the model had trained we would be able to predict on the exact same events just now without any truth values. We would then expect a good model to predict 100% correct or at least close to because the model is so simple. If this isn't the case we would know that there is something wrong with the model or at least that it isn't as good as we want it to be yet. This problem only takes a a short time to run when you have generated your test data compared to the multiple hours it can take to run with large data sets for real data and therefore it is an important step to make before you waste a lot of time training on a bad model.