Pulse shape discrimination at Belle II and μ/π identification

DESY Summer Programme

Noemi Palmeri, University of Rome "La Sapienza", Italy

Supervisor: Savino Longo

September 7, 2021

Abstract

CsI(Tl) pulse shape discrimination (PSD) at the Belle II calorimeter has the potential to improve identification of otherwise indistinguishable particles. In this report, μ/π PSD is implemented by means of a convolutional neural network, which managed to achieve 0.9091 accuracy on a test set by discriminating on pulse shape. Performance on different subsets of pion data was evaluated, isolating ionizing and showering pion clusters. Although showering pions were correctly classified, the network was not able to discriminate ionizing pions from muons. Possible future steps are discussed.

Contents

1 The Belle II experiment

The Belle II experiment, located at the SuperKEKB e^+e^- collider in Tsukuba, Japan, is specifically aimed at achieving a high B-mesons production rate, granting it the nickname of B-factory. For this purpose, electrons and positrons are collided at a center of mass energy of $\sqrt{s} = 10\,580\,\text{MeV}$ to exploit the $\Upsilon(4S)$ [1], which is very likely ($\Gamma_{B\bar{B}}/\Gamma > 96\%$) to decay into a $B\bar{B}$ pair.

1.1 ECL

Among the various Belle II sub-detectors surrounding the beam pipe lies the electromagnetic calorimeter (ECL). It features an array of 8736 CsI(Tl) crystals [2], each capable of emitting scintillation light in response to energy deposition, which is then turned into an electric signal by means of two photodiodes. The shape of the output waveform is related to the total energy deposited in the crystal, E_{crystal} (which is estimated online for all waveforms through a FPGA), proportional to the total scintillation emission, and on the dE/dx of the crossing particle, proportional to the signal's decay constant [3].

The signal produced by the photodiode is processed through several shaping stages, including an integrator (making the final waveform's *peak* proportional to E_{crystal}), the last of which is a 18-bit precision digitizer. 31 points are saved for each waveform, as shown in an example signal in Figure (1). Due to bandwidth limitations, waveforms are only recorded for offline processing if $E_{\text{crystal}} > 50\,\text{MeV}$.

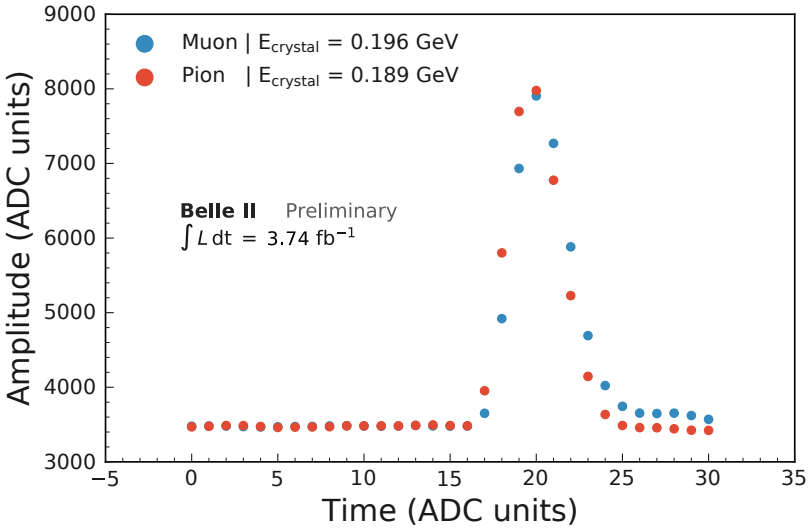


Figure 1: Waveform from Belle II data associated to a muon ($E_{\text{crystal}} = 196\,\text{MeV}$) and a pion ($E_{\text{crystal}} = 189\,\text{MeV}$). For the purpose of comparison, waveforms have been vertically shifted to make the average calculated over the first 15 points equal. The pion waveform decays noticeably faster than the muon's, despite having a similar peak.

Whenever a single particle deposits energy in more than one crystals while travelling through the ECL, several waveforms may be stored.

2 Pulse shape discrimination

It has been experimentally observed that the shape of the waveform generated by CsI(Tl) crystals varies depending on the physical process underlying energy deposition. Specifically, each signal can be decomposed into the sum of a photon-like component, due to low dE/dx ionization, and a hadron-like component, due to high dE/dx ionization (which characterizes hadron signatures) [3]. This feature opens up the possibility of identifying hadrons by analyzing the pulse shape they produce.

At Belle II, pulse shape discrimination is currently implemented through a multi-template offline fit on saved waveforms only, from which two quantities are estimated: total energy deposited in the crystal (`OfflineEnergy`) and energy due to the hadron-like component (`OfflineHadronEnergy`). From these, the fraction of scintillation emission induced by the hadron component relative to total emission is calculated as the ratio between the two (`OfflineHadronIntensity`).

From these premises, non-hadron particles are expected to have a hadron intensity of 0, in contrast to hadrons, for which values strictly above 0 are expected. Imposing a cut threshold on this value would thus ideally allow us to distinguish between the two.

In practice, this technique's effectiveness is limited by the shortcomings of the fit process, such as susceptibility to noise [4]. For this reason, this project attempts to probe an alternative approach based on the use of neural networks.

2.1 Applications

Pulse shape discrimination constitutes a vital tool in the distinction of otherwise almost identical-looking particles, which would prove significant for a number of physical processes. A case in point is the $e^+e^- \to \pi^+\pi^-\gamma$ channel, whose experimentally measured cross-section can allow for calculations of quantum chromodynamics (QCD) corrections to the muon's anomalous magnetic moment [1]. However, a significant background is produced by the $e^+e^- \to \mu^+\mu^-\gamma$ channel: this raises a reason of interest towards μ/π discrimination, which is what the project will focus and try to improve performance on.

3 μ/π identification

Discrimination of μ/π was attempted through the use of a neural network trained on the task of correctly tagging the input samples as pion-like waveforms (Network output = 0) or muon-like waveforms (Network output = 1). Particular efforts were spent to make sure that the network was learning pulse shape differences.

3.1 Data selection

The network was trained on real data produced SuperKEKB collisions. In contrast to simulation data, this poses a potential challenge in terms of sample purity: specific reactions have to be chosen in order to ensure that pure control samples are used.

Muons were extracted by identifying $e^+e^- \to \mu^+\mu^-\gamma$ events. More specifically, $m_{\mu^+\mu^-\gamma}$ (center-of-mass energy) was imposed to be compatible with the collider's $\sqrt{s}$. Once all eligible muon pairs were selected, samples were extracted using a tag and probe system: using the `muonID` variable, which aggregates information from all detectors to estimate the likelihood of a particle actually being a muon, one of the two tracks was identified as a $\mu^\pm$. The other track is then chosen as a sample after verifying that it actually hit the calorimeter.

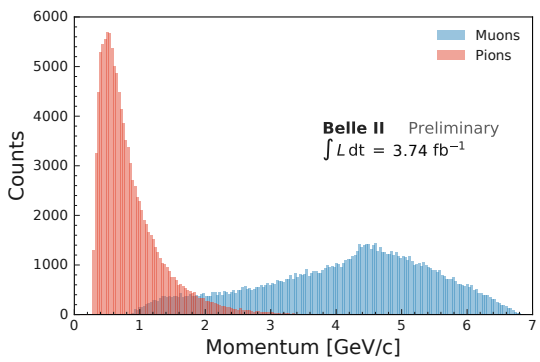
The same procedure is applied to pions, for which the $K^0_S \to \pi^+\pi^-$ decay was employed, using K^0_Ss produced in the $e^+e^- \to K^0_S X$ reaction, with X representing any other possible product. In this case, $m_{\pi^+\pi^-}$ was imposed to be compatible with $m_{K^0_S}$; furthermore, the summed momentum was to point towards the collision vertex. The decay vertex was imposed to be at a distance greater than 0.5 cm from the collider's interaction point. For pions, the equivalent `pionID` variable was used on the tag.

3.2 Data analysis and dataset construction

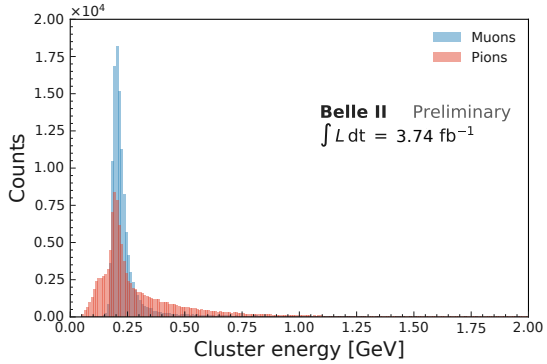
After isolating muons and pions, further processing on the dataset is needed to ensure that physical quantities contributing to dE/dx are similarly distributed, with the sole exception of hadron intensity. This can prevent the network from basing classification on peculiarities of the specific production processes chosen for this analysis, but which cannot be generalized to the entire muon (pion) population. Only discrimination on pulse shape (and, thus, hadron intensity) can translate to better general identification ability.

The distribution of momentum, cluster energy, hadron intensity and crystal energy is shown in Figure (2). For both muons and pions, cluster energy presents a peak at $E_{\text{cluster}} \approx 200\,\text{MeV}$, which is equal to the average energy deposited through ionization by minimum ionizing particles through 30 cm of material (equal to the CsI(Tl) crystal length). For the pions, this accounts for the fraction of particles which manage to cross the calorimeter without showering. In the pion distribution we can also observe an overlapped, exponentially-decaying distribution. This is produced by the pions undergoing hadronic showering [3].

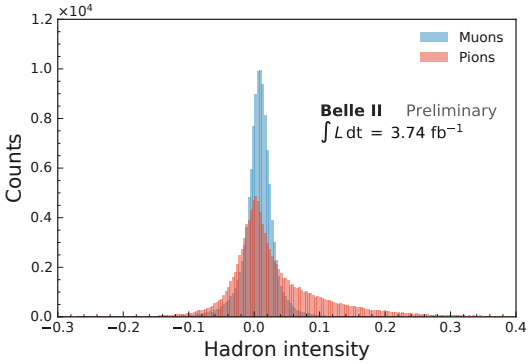
Figure (2c) shows the hadron intensity distribution for the two particles. Though centered around 0 (as expected), the muons' distribution is not perfectly peaking about its mean. As mentioned, this can be attributed to imperfections in the fit process. The pions' distribution, instead, while still centered around 0, features much longer tails, which account for showering particles. To better study the hadron intensity distribution, whose resolution is strongly dependent on crystal energy, a 2D histogram was plotted in Figure (2d), where showering pions can be distinctly seen at higher hadron intensity values.



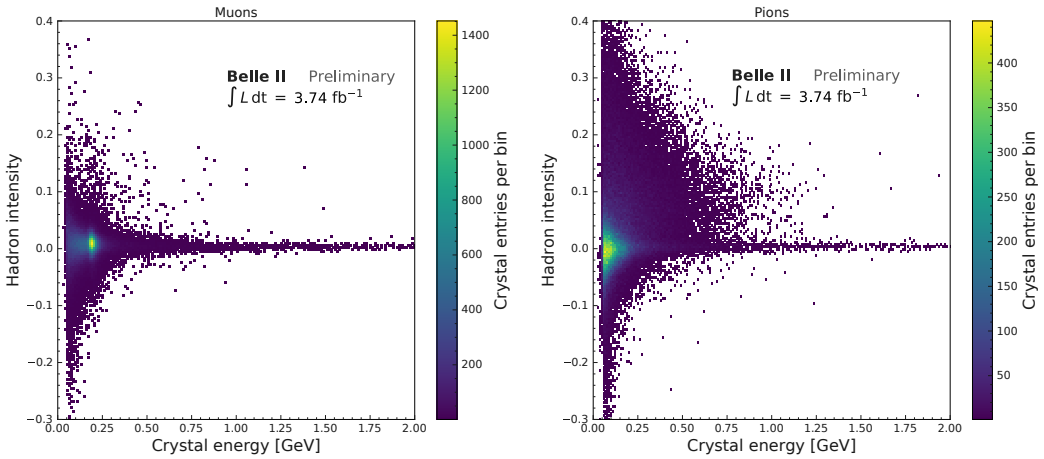
(a) Momentum



(b) Cluster energy



(c) Hadron intensity



(d) Crystal energy vs hadron intensity

Figure 2: Distribution of relevant physical properties for μ/π in the initial dataset.

Hadron intensity is a quantity related to pulse shape. Due to the elevated number of pions with hadron intensity close to 0, it is desirable to remove them to maximize the pulse shape difference between the two particles in the training sample. As such, a Hadron intensity > 0.08 cut was placed.

For the previously stated reason, a cut has been placed on variables directly influencing dE/dx.

For momentum, a $1\,\text{GeV} < p < 2.8\,\text{GeV}$ cut was placed, as it corresponds to the overlapping range of the two histograms.

For crystal energy, instead, the imposed cut was $50\,\text{MeV} < E_{\text{crystal}} < 500\,\text{MeV}$. In this case, however, it is not sufficient to enforce the same range, as the network could pick up on differences in the two particles' distributions. Consequently, re-weighting has been applied on the pions using the crystal energy distribution ratio to replicate the muons' distribution.

To summarize, the following cuts were applied:

- $1\,\text{GeV} < p < 2.8\,\text{GeV}$
- Hadron intensity > 0.08
- $50\,\text{MeV} < E_{\text{crystal}} < 500\,\text{MeV}$
- Re-weighting on cluster energy distribution ratio.

It is important to note that, after re-weighting (see Figure (3)), the final number of waveform entries for muons and pions was roughly equal (~ 19600). For this reason, no further class balancing was needed.

3.3 Neural network

The entire network was implemented using the `Keras` framework with `TensorFlow` as backend [5].

3.3.1 Architecture

Several architectures were tried and trained, varying the type and number of layers. The highest accuracy was achieved using a convolutional neural network (CNN) with the following structure:

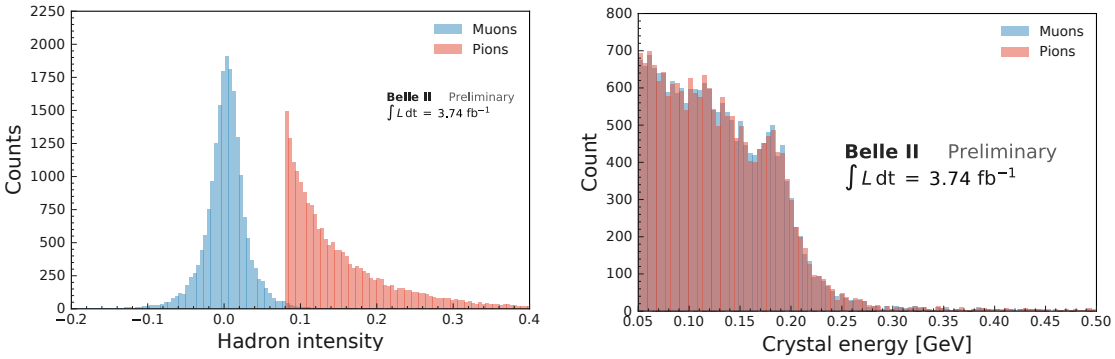
- `Conv1D`, 32 filters, 7 kernel size, tanh activation function;
- `Conv1D`, 64 filters, 5 kernel size, tanh activation function;
- `Dense`, 64 neurons, ReLU activation;
- `Dropout`, 0.2 rate;
- `Dense`, 64 neurons, ReLU activation;
- Output: `Dense`, 1 neuron, sigmoid activation;

By introducing convolutional layers with different kernel sizes, different features can be extrapolated.

In the tests conducted, it was found that increasing network complexity was associated with only a small increase in performance (1–2% accuracy improvement over baseline, 2-layer deep neural network).

3.3.2 Training

After applying the aforementioned cuts on the dataset, as a consequence of which we obtain the crystal energy and hadron intensity distributions shown in Figure (3), only waveforms were extracted to constitute the input samples. Furthermore, to improve network performance, inputs were normalized in the $[0, 1]$ range by dividing each point by $2^{18} - 1$ (maximum ADC count).



(a) Hadron intensity

(b) Crystal energy

Figure 3: Distribution of (a) hadron intensity and (b) crystal energy for μ/π in the training and test set.

The dataset was then split into training (80%) and test set (20%). From the training set, a further 20% was extracted as validation data.

The network was trained using binary crossentropy as loss, with accuracy being tracked as an additional metric. `Adam` was used as optimizer, with a starting learning rate of 0.001 and an implemented callback that further reduce learning rate down to 10^{-6} whenever loss improved by less than 0.005 in 100 epochs.

Training proceeded for 1500 epochs with mini-batches of 32 `batch_size`. Loss and accuracy on the training and validation set are plotted in Figure (4). No overfitting can be observed. Loss plateaued after about 800 epochs.

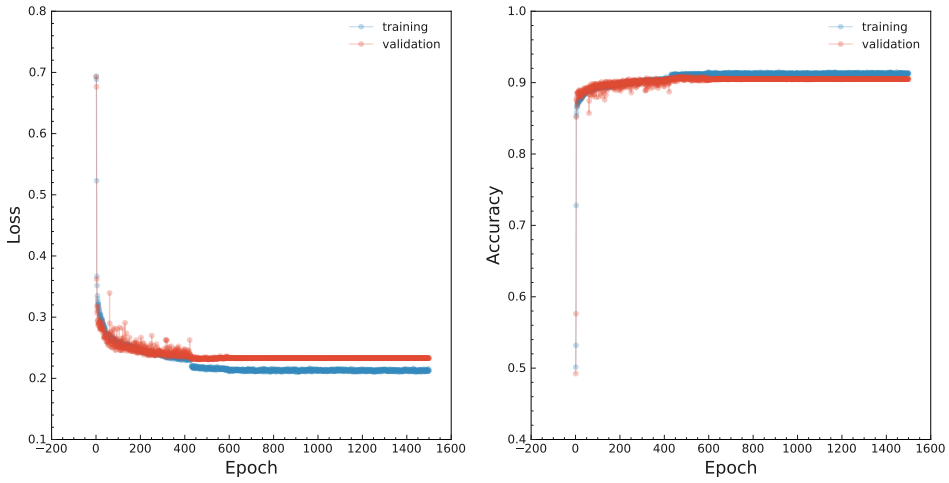


Figure 4: Loss and accuracy per training epoch for training and validation set.

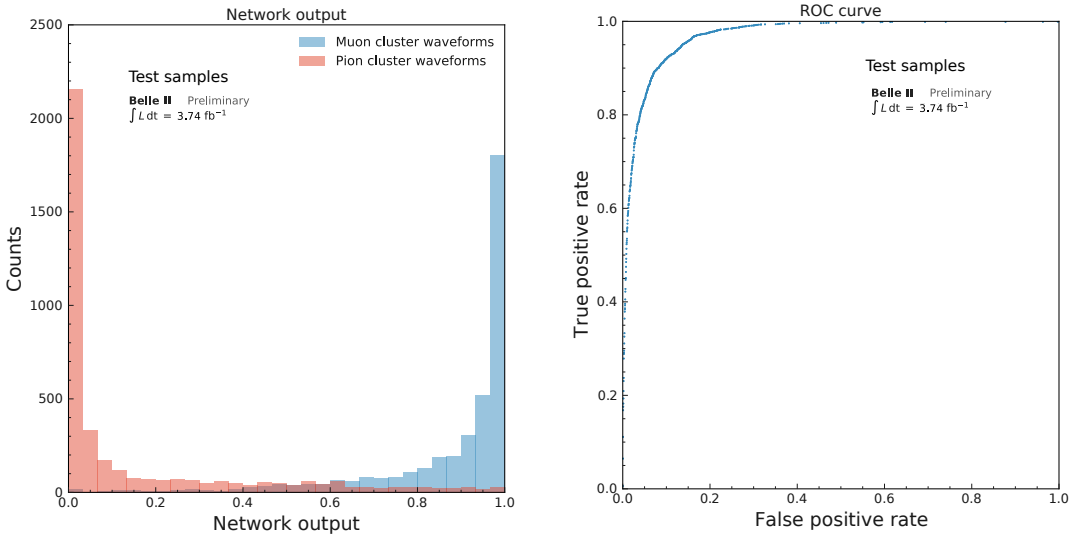
After training, the best model evaluated on validation accuracy was loaded, which scored 0.9090 accuracy.

3.3.3 Evaluation

To properly attest the network's performance, accuracy and AUC (Area Under the ROC Curve) was estimated on the test set. Network output and ROC curve are reported in Figure (5).

The reported scores are:

$$\text{Accuracy} = 0.9091 \qquad \text{AUC} = 0.9697$$



(a) Network output **(b)** ROC

Figure 5: (a) Network output and (b) ROC curve evaluated on test set. AUC = 0.9697.

Both indicators, as it can be easily assessed from the network output separation, indicate a good separation between pions and muons. Furthermore, as a sanity check on training, test accuracy is comparable to validation accuracy.

3.4 Inference

As highlighted by Figure (2), two different pion populations can be distinguished: ionizing pions, which are the most similar to muons, and showering pions, which feature a high hadron intensity value and are thus the easiest to tell apart. After training the model on a dataset crafted to maximize pulse shape difference, it can be useful to assess the network's discrimination capability on specific subsets of samples.

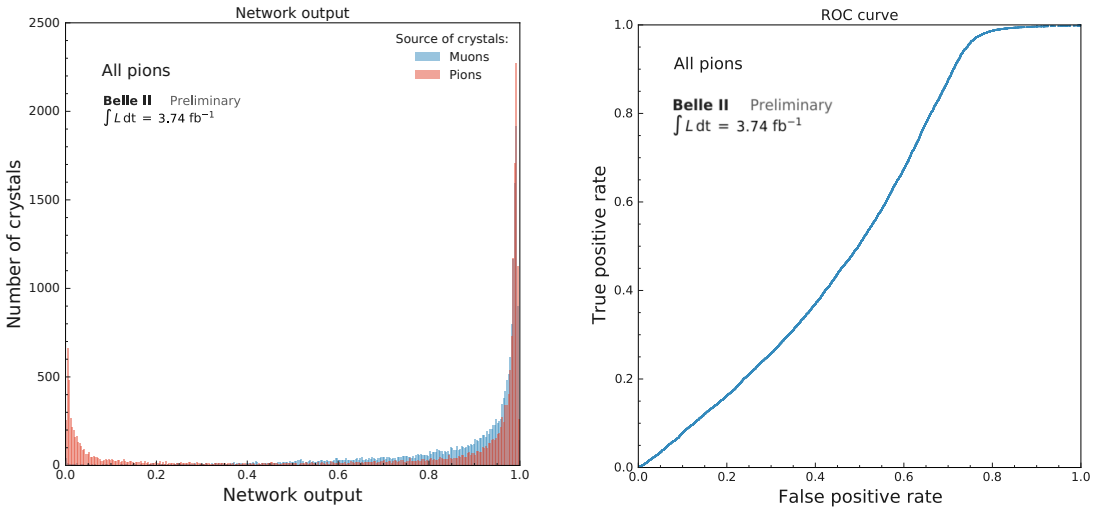
In order to still provide the network with comparable samples, the p cut was maintained on both the muon and pion dataset. However, whereas the muon

dataset was kept the same for all ensuing trials, different cuts were applied to the pion one to isolate ionizing and showering particles.

Whenever one of the two datasets was found to outnumber the other, classes were rebalanced through undersampling.

3.4.1 Whole dataset

As a first evaluation, the network was asked to infer output on the entire dataset, without further cuts applied. Output and ROC curve is represented in Figure (9).



(a) Network output

(b) ROC

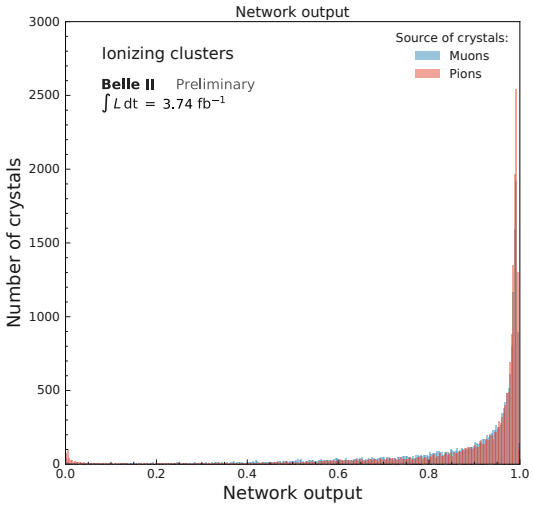
Figure 6: (a) Network output and (b) ROC curve evaluated on crystals from the entire dataset. AUC = 0.5415.

While the two peaks around 0 and 1 can be still observed, it is evident that the majority of pion waveforms was misclassified as a muon. By isolating the ionizing and showering pions contribution, we hope to be able to explain the two different contributions to the plot.

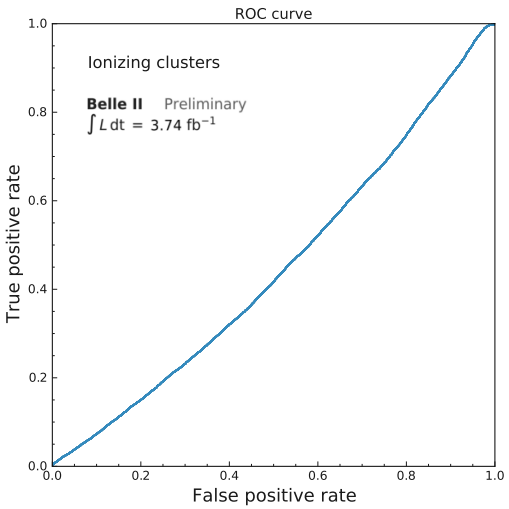
3.4.2 Ionizing pions

In order to isolate ionizing pions, a $150\,\text{MeV} < E_{\text{cluster}} < 300\,\text{MeV}$ was placed. Furthermore, for reasons similar to the E_{crystal} case for the training set, re-sampling was used on the momentum distribution ratio. Network output and ROC curve are shown in Figure (7a).

As expected, the network output distributions are almost completely overlapped. A curious observation is that AUC < 0.5, which implies that the classifier is actively misclassifying most pulses. This should not come as a surprise, as the dE/dx values of ionizing pions are almost completely identical to muons'.



(a) Network output



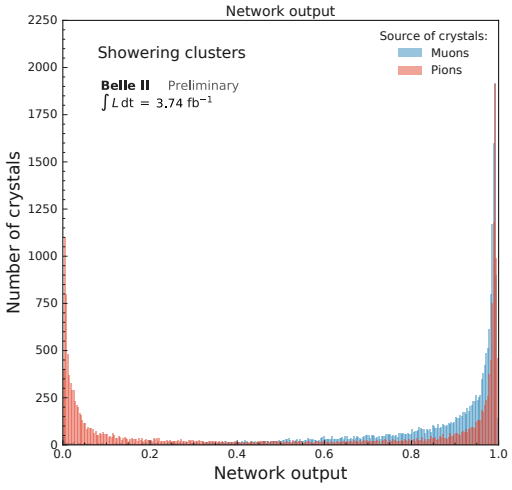
(b) ROC

Figure 7: (a) Network output and (b) ROC curve evaluated on crystals from the entire muon dataset against ionizing pions only. AUC = 0.4480.

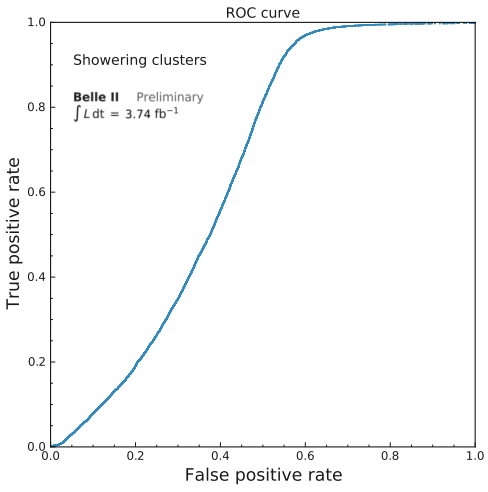
3.4.3 Showering pions

To isolate showering pions, it is sufficient to invert the E_{cluster} cut imposed in the ionizing pions case ($E_{\text{crystal}} < 150\,\text{MeV}$ or $E_{\text{crystal}} > 300\,\text{MeV}$). Network output and ROC curve are shown in Figure (8).

Both plots are similar to their counterpart for the whole dataset. However, a higher AUC score is achieved thanks to the absence of ionizing pions.



(a) Network output



(b) ROC

Figure 8: (a) Network output and (b) ROC curve evaluated on crystals from the entire muon dataset against showering pions only. AUC = 0.6442.

It should be pointed out that, for these trials, cuts were imposed on cluster variables, rather than crystal ones. The distinction made, based on a cluster energy threshold, is thus between showering clusters and ionizing clusters. In the former category, the selection therefore includes all crystals ionized by a pion before it reacted hadronically with the material's nuclei within the same cluster, and whose presence justifies the peak at Network output = 1. The number of such pions is consistent with what is expected by observing that the CsI(Tl) nuclear interaction length $X_0 = 38.04\,\text{cm}$ [6] is greater than crystal length $L_{\text{crystal}} = 30\,\text{cm}$.

4 Conclusions

4.1 Results

A convolutional network was trained on the task of classifying input waveforms, with proper cuts imposed on data to ensure pulse shape difference between the two classes, as pion-like (0) or muon-like (1). Training proceeded as planned, achieving 0.9091 classification accuracy on the test set.

The network was tested on the initial dataset to probe accuracy on real data, imposing different cuts in order to analyze specific scenarios. More specifically, the network was run on ionizing and showering pions, obtaining different performance. In particular, ionizing pions were not distinguished from ionizing muons; showering pions, instead, showed much more separation, accounting for most of the correct classifications when the network was run on the complete dataset.

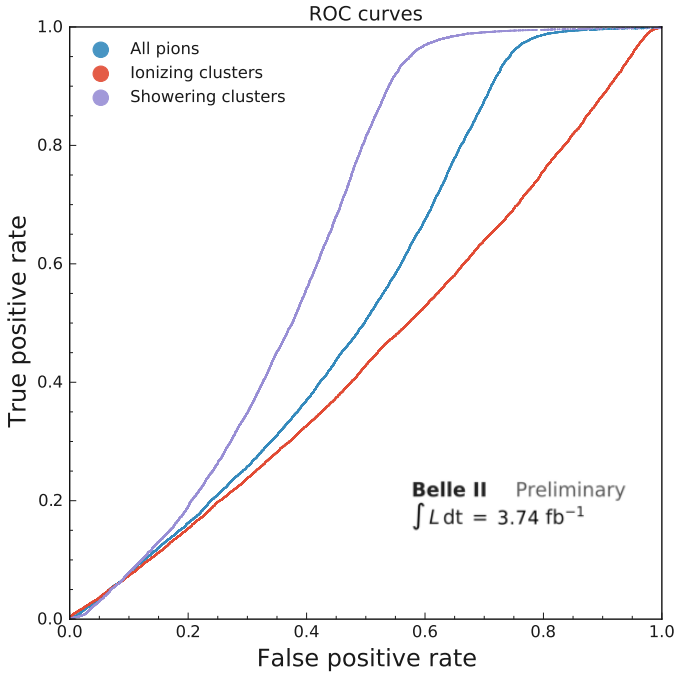


Figure 9: Network output and ROC curve evaluated on crystals from the entire muon dataset against different pion datasets.

	Whole	Ionizing	Showering
AUC	0.5415	0.4480	0.6442

Table 1: AUC scores for the three different pion datasets used for network inference. Muon dataset was kept the same for all three trials.

4.2 Future steps

There are several directions the proposed approach could be improved in, as well as different points still left to better investigate.

- **Hyperparameter tuning and improved training** The results presented in this report can be potentially still be improved upon by conducting a finer hyperparameter tuning. Furthermore, as several difficulties was encountered in training more complex architectures, such as vanishing gradient, it is possible that deeper networks, if properly trained, could achieve higher accuracy.

- **Ionizing pions and muons** Ionizing pions were incorrectly classified as muons. The similar mass of the two particles, combined with the fact that energy deposition is driven by the same physical process, could make them indistinguishable, should the detector's resolution prove insufficient for the task.

  In order to exclude this eventuality, further training on the ionizing pions sample is needed. Some preliminary tests conducted with several architectures showed discouraging results, with accuracy close to 50%. However, further runs should be conducted in order to rule out training issues.

  Should the task prove unfeasible, a plausible explanation for this impossibility can be hypothesized by taking into account the variations of crystal response in the calorimeter. Each of the CsI(Tl) crystals, provided with the same input (energy deposit and corresponding physical process), outputs a slightly different signal, increasing the uncertainties on the waveform. When considering that the sample of waveforms used is put together from thousands of different crystals, such uncertainties could prove to be too high to resolve the slight differences in pulse shape between ionizing muons and pions, leading to impossible discrimination.

- **Different approaches to training** In practical applications, the correct task to pursue is to correctly identify particles on a per-event basis, rather than per-crystal. Although this would slightly deviate from the scope of the current project, this could result in enhanced discrimination capability. Consequently, it could prove beneficial to use all waveforms belonging to the same cluster as individual input samples, rather than single pulses.

  Some preliminary tests conducted on DNNs, in which input samples were crafted by concatenating all waveforms produced by the same particle for each event, showed promising results, with accuracy close to 70% when

trained on ionizing pions only (though it emerged from post-training analysis that such a high accuracy score could be due to pecularities introduced in the training set by dataset processing). However, such attempts do not in any way represent neural networks' actual discriminating power in the ionizing muons/pions set, and, as such, further investigation is needed.

References

[1] E Kou et al. "The Belle II Physics Book". In: *Progress of Theoretical and Experimental Physics* 2019.12 (Dec. 2019). ISSN: 2050-3911. DOI: 10.1093/ptep/ptz106. URL: http://dx.doi.org/10.1093/ptep/ptz106.

[2] T. Abe et al. *Belle II Technical Design Report*. 2010. arXiv: 1011.0352 [physics.ins-det].

[3] S. Longo et al. "CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron–positron colliders". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 982 (2020), p. 164562. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.164562. URL: https://www.sciencedirect.com/science/article/pii/S0168900220309591.

[4] Stella Katharina Wermuth, Dr. Torben Ferber, and Prof. Dr. Oliver Gerberding. "Neural Network based Pulse Shape Analysis with the Belle II Electromagnetic Calorimeter". Presented on 22 06 2021. PhD thesis. Hamburg: Hamburg, Universität Hamburg, 2021.

[5] Noemi Palmeri. *CNN for Pulse Shape Discrimination*. https://github.com/noepalm/cnn-psd. 2021.

[6] Particle Data Group. *Atomic and nuclear properties of materials: CsI*. https://pdg.lbl.gov/2012/AtomicNuclearProperties/HTML_PAGES/141.html.