

# Improved calorimetric particle identification in NA62 using machine learning techniques

---

## The NA62 collaboration

*E-mail:* [na62eb@cern.ch](mailto:na62eb@cern.ch)

**ABSTRACT:** Measurement of the ultra-rare  $K^+ \rightarrow \pi^+ \nu \bar{\nu}$  decay at the NA62 experiment at CERN requires high-performance particle identification to distinguish muons from pions. Calorimetric identification currently in use, based on a boosted decision tree algorithm, achieves a muon misidentification probability of  $1.2 \times 10^{-5}$  for a pion identification efficiency of 75% in the momentum range of 15–40 GeV/ $c$ . In this work, calorimetric identification performance is improved by developing an algorithm based on a convolutional neural network classifier augmented by a filter. Muon misidentification probability is reduced by a factor of six with respect to the current value for a fixed pion-identification efficiency of 75%. Alternatively, pion identification efficiency is improved from 72% to 91% for a fixed muon misidentification probability of  $10^{-5}$ .

**KEYWORDS:** Fixed Target Experiments, Branching fraction, Rare Decay, Flavour Physics

**ARXIV EPRINT:** [2304.10580](https://arxiv.org/abs/2304.10580)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>NA62 beam and detector</b>	<b>1</b>
<b>3</b>	<b>GBM and CNN algorithms</b>	<b>3</b>
<b>4</b>	<b>Data samples</b>	<b>4</b>
<b>5</b>	<b>Particle identification algorithm</b>	<b>5</b>
<b>6</b>	<b>Performance of the algorithm</b>	<b>7</b>
<b>7</b>	<b>Summary</b>	<b>8</b>
	<b>The NA62 collaboration</b>	<b>11</b>

---

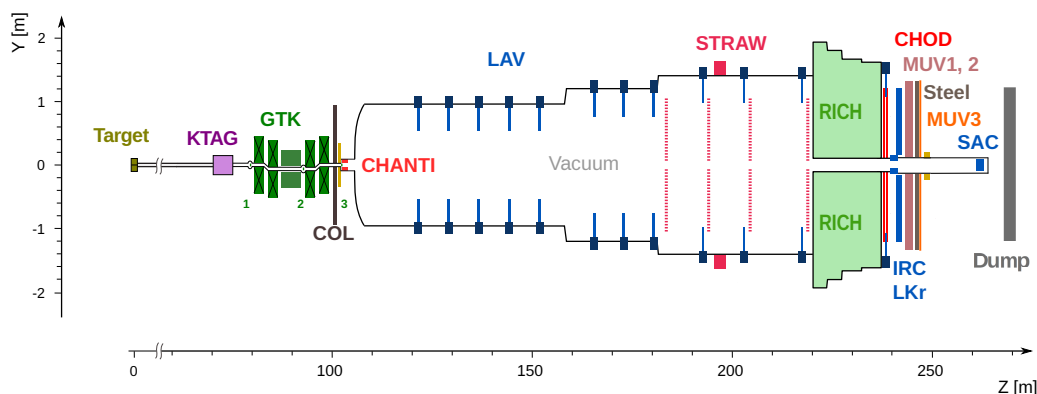
## 1 Introduction

The NA62 experiment at CERN is dedicated to measurements of charged kaon decays, including the highly-suppressed “golden mode”  $K^+ \rightarrow \pi^+ \nu \bar{\nu}$  [1] with a Standard Model branching ratio of  $(8.4 \pm 1.0) \times 10^{-11}$  [2]. Suppression of the  $K^+ \rightarrow \mu^+ \nu$  background to the  $K^+ \rightarrow \pi^+ \nu \bar{\nu}$  decay requires a  $\mu^+$  misidentification probability below  $10^{-7}$ , which is achieved by a combination of calorimetric  $\pi^+$  identification providing a muon suppression factor of  $\mathcal{O}(10^{-5})$ , and Cherenkov  $\pi^+$  identification providing an additional suppression factor of  $\mathcal{O}(10^{-2})$ .

Machine learning (ML) methods developed primarily for image recognition can be applied to particle identification (PID) at NA62. Convolutional neural networks (CNN), a type of deep neural network, have been successfully used for classification tasks commonly encountered in particle physics [3–5]. In this work, a calorimetric PID algorithm based on ML models is designed for the NA62 experiment. Two ML models are considered: LightGBM [6], which is an implementation of a gradient boosting machine (GBM) [7], and a CNN-based model. Performance of the algorithm is compared to that of the boosted decision tree (BDT) implemented using TMVA GradientBoost [8], currently used by NA62.

## 2 NA62 beam and detector

The layout of the NA62 detector [9] is shown schematically in figure 1. An unseparated secondary beam of  $\pi^+$  (70%), protons (23%), and  $K^+$  (6%) is created by directing 400 GeV/ $c$  protons extracted from the CERN SPS onto a beryllium target in spills of 3 s effective



**Figure 1.** Schematic side view of the NA62 detector used in 2018.

duration. The central beam momentum is  $75 \text{ GeV}/c$ , and the momentum spread is 1% (rms). Beam kaons are tagged with 70 ps time resolution by a differential Cherenkov counter (KTAG) using a nitrogen gas radiator at 1.75 bar pressure contained in a 5 m long vessel. Beam particle positions, momenta and times are measured by a silicon pixel spectrometer consisting of three stations (GTK1, 2, 3) and four dipole magnets. A 1.2 m thick steel collimator (COL) is placed upstream of GTK3 to absorb hadrons produced in upstream  $K^+$  decays. Inelastic interactions of beam particles in the GTK are detected by an array of scintillator hodoscopes (CHANTI). The beam is delivered into a vacuum tank evacuated to a pressure of  $10^{-6}$  mbar, which contains a 75 m long fiducial decay volume (FV) starting 2.6 m downstream of GTK3.

Downstream of the FV, undecayed beam particles continue their path in vacuum. Momenta of charged particles produced in  $K^+$  decays in the FV are measured by a magnetic spectrometer (STRAW) located in the vacuum tank downstream of the FV. The spectrometer consists of four tracking chambers made of straw tubes, and a dipole magnet located between the second and third chambers that provides a horizontal momentum kick of  $270 \text{ MeV}/c$ . The spectrometer momentum resolution is  $\sigma_p/p = (0.30 \oplus 0.005 \cdot p)\%$ , where the momentum  $p$  is expressed in  $\text{GeV}/c$ .

A ring-imaging Cherenkov detector (RICH), consisting of a 17.5 m long vessel filled with neon at atmospheric pressure (with a Cherenkov threshold for pions of  $12.5 \text{ GeV}/c$ ), is used for the identification of charged particles and for time measurement with 70 ps precision. Two scintillator hodoscopes (CHOD) located downstream of the RICH provide trigger signals and time measurements with 200 ps precision.

A quasi-homogeneous liquid krypton (LKr) electromagnetic calorimeter with a thickness equivalent to 27 radiation lengths or 2.1 nuclear interaction lengths is used for particle identification and photon detection. The calorimeter has an active volume of  $7 \text{ m}^3$ , is segmented in the transverse plane into 13248 projective cells of  $2 \times 2 \text{ cm}^2$ , and provides an energy resolution  $\sigma_E/E = (4.8/\sqrt{E} \oplus 11/E \oplus 0.9)\%$ , where  $E$  is expressed in GeV. To achieve hermetic acceptance for photons emitted in the FV by  $K^+$  decays at angles up to 50 mrad to the beam axis, the LKr calorimeter is supplemented by annular lead glass

detectors (LAV) installed in 12 positions inside and downstream of the vacuum tank, and two lead/scintillator sampling calorimeters (IRC, SAC) located close to the beam axis.

A steel/scintillator hadronic sampling calorimeter is formed of two modules (MUV1, 2) with transverse dimensions of approximately  $2.6 \times 2.6 \text{ m}^2$ . The MUV1 (MUV2) module consists of 24 steel plates of 26.8 (25) mm thickness, interleaved with 9 (4.5) mm thick scintillator layers, resulting in a total thickness of 4.1 (3.7) interaction lengths. The scintillator layers in both modules are made of strips aligned alternately in the horizontal and vertical directions in consecutive planes. Each MUV1 scintillator layer consists of 48 strips, including 40 strips spanning the entire transverse length of the detector (read out at both ends), and 8 central strips of half-length (read out at the peripheral end), leading to  $2 \times 44$  channels per plane. Each MUV2 scintillator layer consists of  $2 \times 22$  strips, each spanning half the transverse length and read out at the peripheral end. In both modules, all strips located in the same transverse position in different layers are coupled to the same photomultiplier, thus providing a horizontal and a vertical view. A muon detector (MUV3) is located behind a 80 cm (4.8 interaction lengths) thick iron wall, has a transverse size of  $2.64 \times 2.64 \text{ m}^2$ , and is built from 50 mm thick scintillator tiles, including 140 regular tiles of  $220 \times 220 \text{ mm}^2$  transverse dimensions, and eight smaller central tiles adjacent to the beam pipe.

A two-level trigger system [10, 11] is employed to reduce the event rate from 10 MHz to about 100 kHz. Multiple trigger lines with different downscaling factors are operated concurrently.

### 3 GBM and CNN algorithms

Supervised ML algorithms are used to construct a function (called *model* or *classifier*),  $f(x_i; \theta)$ , mapping a set of inputs  $x_i$  (called *examples*) consisting of a number of attributes (called *features*) to a set of outputs  $y_i$  (called *labels*). The set of model parameters,  $\theta$ , is chosen to minimise a *loss function* (also called *objective function*),  $\mathcal{L} = \sum_i \ell(y_i, f(x_i; \theta))$ , where the function  $\ell(y, f(x; \theta))$  is selected depending on the problem to solve [12, 13].

GBM is a generic family of algorithms used to build strongly predictive classifiers as linear combinations of *weak* classifiers (such as decision trees). The classifier,  $f(x_i)$ , is constructed iteratively, starting from an initial decision tree. At each step  $m$ , residuals are computed using the loss function obtained at the previous step as  $-\partial \ell(y_i, f_{m-1}(x_i)) / \partial f_{m-1}$ , and a regression tree fitting the residuals is constructed and added to the linear combination. This procedure is repeated until a stopping condition is met [7].

CNN algorithms have been developed to address computer vision tasks such as translation-invariant image classification, and can recognize complex geometric features at multiple scales. CNNs consist of consecutive *layers*, the output of each layer being the input to the next one. Each layer is represented by a function  $g(x)$ , where  $x$  is an input tensor. Common layer types include convolutional, batch normalisation and pooling layers [12, 13]. It has been observed [14] that increasing the number of layers leads to degradation of the training performance; to mitigate this issue, the ResNet architecture [14] introduces groups of consecutive layers called the *residual blocks*. Instead of learning the parameters of the

underlying function  $g(x)$  directly, the block learns the parameters of the residual function  $g_r(x)$ , encoding the difference between the input and output of the layer. Many of these blocks are typically stacked.

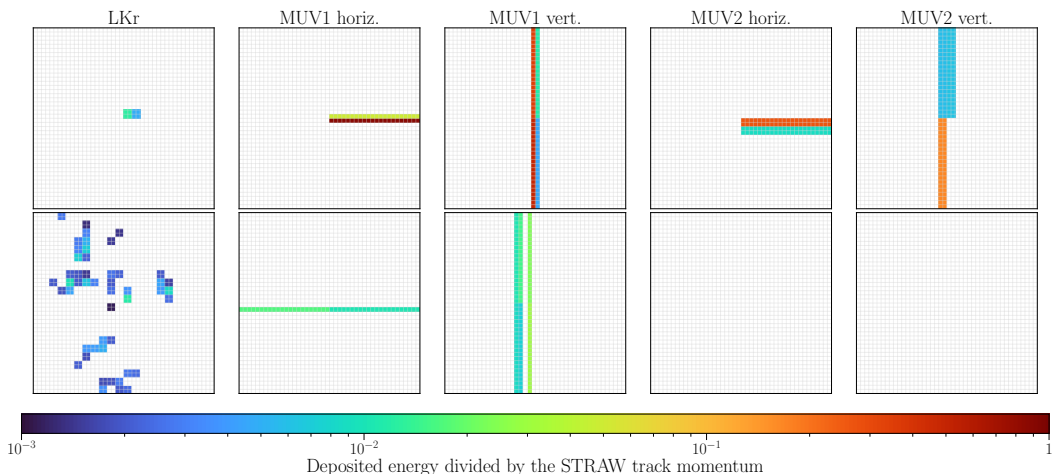
## 4 Data samples

Two independent datasets are employed: a *training and validation dataset* is used to train the models and adjust the hyperparameters (i.e. the parameters not derived via training), while an independent *test dataset* is used to characterise the performance of the optimal model with a frozen set of hyperparameters. The training and validation (test) dataset is based on 2.5% (1.5%) of the data sample collected by the NA62 experiment in 2016–2018. The training and validation dataset contains events collected with a non-downscaled  $K^+ \rightarrow \pi^+\nu\bar{\nu}$  trigger line [1] based on RICH and CHOD signals in the absence of in-time MUV3 signals, and a control trigger line based on the CHOD signal downscaled by a factor of 400. The test dataset contains events collected with the control trigger line. This provides sufficient statistics for training, ensuring that the test dataset is free from trigger-induced bias.

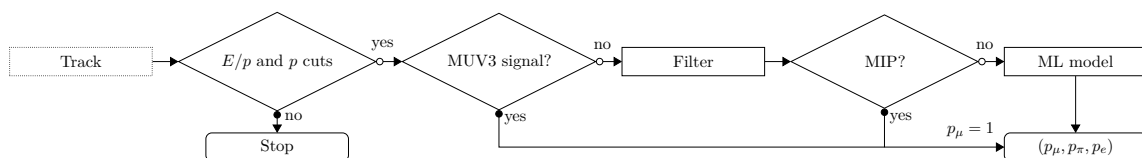
High-purity pion ( $\pi^+$ ), muon ( $\mu^+$ ) and positron ( $e^+$ ) track samples in the momentum range 15–50 GeV/ $c$  are obtained from  $K^+ \rightarrow \pi^+\pi^0$ ,  $K^+ \rightarrow \mu^+\nu$  and  $K^+ \rightarrow \pi^0e^+\nu$  decay candidates selected without calorimetric particle identification, following the procedure used earlier for the existing NA62 BDT algorithm [15]. Isolated STRAW tracks with associated in-time KTAG, CHOD, RICH and LKr signals and a single associated GTK track are considered. The decay vertex, reconstructed as the point of closest approach of the STRAW and GTK tracks, is required to be located in the FV and within the beam envelope. For the  $K^+ \rightarrow \pi^+\pi^0$  and  $K^+ \rightarrow \pi^0e^+\nu$  decays, a prompt  $\pi^0 \rightarrow \gamma\gamma$  decay is reconstructed by measuring the photons in the LKr calorimeter. Further selection criteria are based on photon veto conditions, particle identification in the RICH, and missing mass squared,  $m_{\text{miss}}^2 = (P_K - P_{\text{tr}})^2$ , where  $P_K$  and  $P_{\text{tr}}$  are the  $K^+$  and decay track four-momenta, respectively. A label ( $\pi^+$ ,  $\mu^+$  or  $e^+$ ) is assigned to each selected track.

The following features are encoded as matrices and vectors: track momentum and impact position in the LKr front plane; presence of a matching MUV3 signal; energy deposits in a matrix of  $22 \times 22$  LKr calorimeter cells centered around the track impact position (these matrices are sparse due to the LKr Molière radius of 4.7 cm); total LKr energy deposit associated to the track; energy deposits in the horizontal and vertical views in all MUV1 channels ( $2 \times 44$  channels in each view) and all MUV2 channels ( $2 \times 22$  channels in each view). Energy deposits within 10 ns of the track time are considered; the energy deposit values in every channel or cell are divided by the track momentum.

The GBM algorithms require input information in the form of a single vector for each track. Therefore the LKr, MUV1 and MUV2 energy deposit matrices are transformed into vectors; the LKr matrix is cropped to a size of  $18 \times 18$  to exploit its sparsity. The results are concatenated into a 588-element feature vector containing the complete information about the track.



**Figure 2.** CNN input matrices of  $44 \times 44$  size constructed for a  $\mu^+$  track (top row) and a  $\pi^+$  track (bottom row) from the training dataset. The LKr matrix is centred at the track impact point in the LKr front plane. In the MUV matrices related to the horizontal (vertical) view, the vertical (horizontal) coordinate is centered at the track impact point in the LKr front plane.



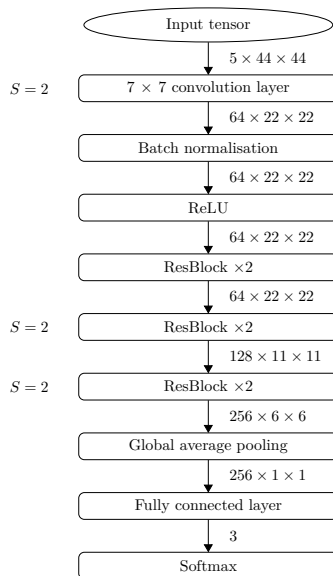
**Figure 3.** Flow chart of the developed calorimetric PID algorithms.

The CNN processing aims to exploit the correlations between the LKr, MUV1 and MUV2 signals. Most CNN architectures require an arbitrary number of matrices of identical dimensions (known in the ML context as *channels*) as input. Therefore the LKr, MUV1 and MUV2 matrices are transformed to a fixed size of  $44 \times 44$  by replicating the element values (though the LKr and MUV matrices correspond to different transverse areas). The resulting matrices are illustrated for two events in figure 2. The inputs are finally arranged into a  $5 \times 44 \times 44$  input tensor.

## 5 Particle identification algorithm

Most muons traversing the calorimeters behave as minimum-ionising particles (MIPs). However, in rare cases, radiative processes such as bremsstrahlung or pair-production occur, resulting in signal patterns that mimic those of charged pions. These cases are identified by the developed algorithms through the shower shape analysis.

The flow of the developed PID algorithm is illustrated in figure 3. The algorithm returns the probabilities  $p_\pi, p_\mu, p_e$  of the track to be classified as  $\pi^+, \mu^+, e^+$ , respectively. Tracks with LKr energy deposit to momentum ratio  $E/p < 0.95$  are considered, which reduces the  $e^+$  contamination in the  $\pi^+$  and  $\mu^+$  samples. Tracks with matching MUV3 signals, or identified as MIPs by a dedicated filter, are assigned  $p_\mu = 1$ . For the remaining tracks, a



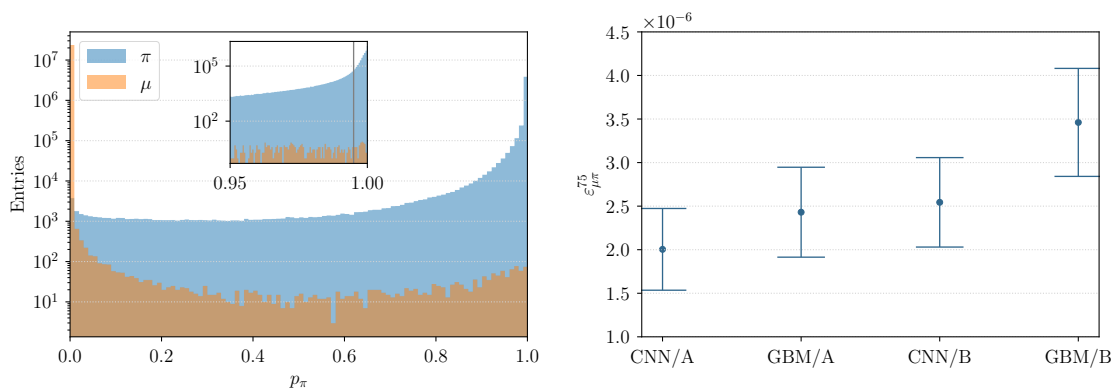
**Figure 4.** The NA62ResNet architecture. Residual blocks (ResBlock) are stacked in pairs. The stride  $S$  [13] is a parameter of the filter. The output tensor dimensions at each step are indicated.

ML model (either a GBM or a CNN algorithm) is employed to evaluate the probabilities  $p_\pi$ ,  $p_\mu$ ,  $p_e$ . The models are trained for tracks in the momentum range of 15–40 GeV/ $c$  used in the  $K^+ \rightarrow \pi^+ \nu \bar{\nu}$  analysis [1].

Two MIP filter algorithms, which use the same input information as the ML models, are considered. The filter A identifies MIPs as tracks with at most four geometrically associated in-time signals in LKr calorimeter cells, at most two hits in each MUV1 view, and at most one hit in each MUV2 view. The filter B additionally exploits the fact that muons typically produce narrow calorimeter signal patterns. Both ML models have been tested with each of the two filters; the corresponding setups are referred to as GBM/A, GBM/B, CNN/A and CNN/B.

The GBM implementation uses the LightGBM framework, which is efficient for large datasets and high-dimensional feature spaces. A CNN architecture of the ResNet-18 type [14] is used. To handle the sparsely populated input matrices (figure 2), the  $3 \times 3$  maximum pooling layer after the first batch normalisation layer is removed. The network is further simplified by removing the last two residual blocks that include four 512-input convolution layers. The output of the last of the remaining residual blocks is down-sampled via a global average pooling layer, which is followed by a fully connected layer producing the final output subsequently normalised using a softmax function to obtain the probabilities  $p_\pi$ ,  $p_\mu$  and  $p_e$ . The resulting NA62ResNet architecture is shown in figure 4.

Data augmentation is used to improve the robustness and generalisation of the CNN model: for each epoch of the training, the five input matrices are mirror-imaged with respect to the horizontal axis for 50% of the tracks in the training sample chosen at random. The weights and the bias terms of the fully connected layer are initialised with random numbers distributed uniformly in the  $(-1/\sqrt{N}, 1/\sqrt{N})$  range, where  $N = 256$  is the number of input



**Figure 5.** Left:  $p_\pi$  distributions for  $\pi^+$  and  $\mu^+$  tracks in the momentum range of 15–40 GeV/ $c$  obtained with the CNN/A setup for the test dataset. Inset: the high- $p_\pi$  region; the threshold  $p_0$  chosen for  $\pi^+$  identification is indicated with a vertical line. Right:  $\mu^+$  misidentification probabilities ( $\varepsilon_{\mu\pi}^{75}$ ), which are strongly correlated, with their statistical uncertainties. The statistical uncertainties are measured for each setup using 500 datasets constructed by sampling the test dataset with replacement.

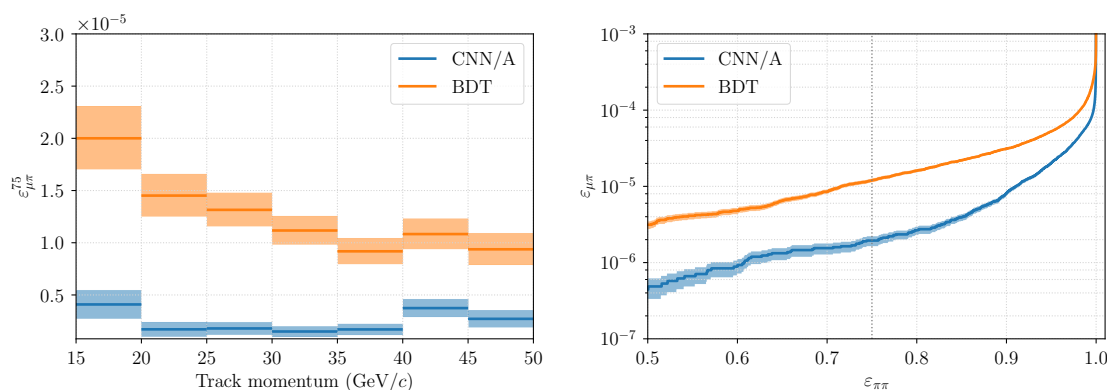
features. The convolution layers are initialised using the He scheme [16], and the cross entropy loss function is used. The network parameters are optimised following the Adam method [17]. To detect potential overfitting, the validation loss is monitored during training. The PyTorch framework [18] is used to build and train the model. The architecture is integrated into the NA62 software framework.

The numbers of  $\pi^+$ ,  $\mu^+$  and  $e^+$  tracks in the training and validation dataset passing the MIP filter A and used for training (figure 3) are  $6.9 \times 10^6$ ,  $1.0 \times 10^7$  and  $7.8 \times 10^4$ , respectively. During the model training, 25% of the dataset is randomly set aside to form a validation sample, and the rest is allocated to the training sample. The cross entropy loss function [13] is used for training of both GBM and CNN models.

## 6 Performance of the algorithm

The results presented in this section are obtained using the test dataset. Background contamination in the  $\pi^+$  and  $\mu^+$  samples in the dataset is found using Monte Carlo simulations to be below  $10^{-5}$  and  $10^{-6}$ , respectively. The  $p_\pi$  distributions for  $\pi^+$  and  $\mu^+$  tracks obtained with the CNN/A setup for tracks in the momentum range of 15–40 GeV/ $c$  are shown in figure 5 (left). A condition  $p_\pi > p_0$  is used for  $\pi^+$  identification, where  $p_0$  is chosen for each setup to obtain a  $\pi^+$  identification efficiency of 75%. The corresponding  $\mu^+$  misidentification probabilities,  $\varepsilon_{\mu\pi}^{75}$ , displayed in figure 5 (right), and pairwise differences in  $\varepsilon_{\mu\pi}^{75}$  among the setups, are measured with the test dataset. The uncertainties on these quantities are measured using 500 datasets constructed by sampling the test dataset with replacement [19]. Pairwise difference tests indicate that the variation in  $\varepsilon_{\mu\pi}^{75}$  between CNN/A and CNN/B, as well as between CNN/A and GBM/B, is statistically significant. The





**Figure 6.** Left: muon misidentification probability ( $\varepsilon_{\mu\pi}^{75}$ ) as a function of track momentum, measured with the CNN/A setup for the test dataset. Right: the ROC curve, i.e. muon misidentification probability ( $\varepsilon_{\mu\pi}$ ) vs pion identification efficiency ( $\varepsilon_{\pi\pi}$ ) for a set of  $p_0$  values, obtained with the CNN/A setup for the test dataset in the 15–40 GeV/ $c$  momentum range. The selected operating point,  $\varepsilon_{\pi\pi} = 0.75$ , is indicated by the vertical dotted line. The performance of the BDT algorithm currently used by NA62 is also shown in both panels. The shaded areas represent the statistical uncertainties.

CNN/A setup is chosen over GBM/A for further analysis as it provides a lower central  $\varepsilon_{\mu\pi}^{75}$  value.

The measured  $\varepsilon_{\mu\pi}^{75}$  value as a function of track momentum is shown in figure 6 (left). Strong  $\mu^+$  suppression is achieved also in the momentum range 40–50 GeV/ $c$  not used for training. The receiver operating characteristic (ROC) curve in the momentum range 15–40 GeV/ $c$  is shown in figure 6 (right). An improvement in  $\varepsilon_{\mu\pi}^{75}$  by a factor of six, from  $1.2 \times 10^{-5}$  to  $2.0 \times 10^{-6}$ , is obtained with respect to the BDT algorithm currently used by NA62 [15]. Alternatively, the  $\pi^+$  identification efficiency in the above momentum range is increased from 72% to 91% when the muon misidentification probability  $\varepsilon_{\mu\pi}$  is kept at a fixed level of  $10^{-5}$ . The  $e^+$  misidentification probability ( $\varepsilon_{e\pi}^{75}$ ) is 35% lower in comparison to the BDT algorithm.

Training the CNN/A model separately in each momentum bin leads to an increase of  $\varepsilon_{\mu\pi}^{75}$  by a factor of two across the momentum bins in comparison with figure 6, which is attributed to the smaller size and  $\pi^+/\mu^+$  imbalance of the training and validation datasets in the individual bins. Training the model in the entire 15–50 GeV/ $c$  range increases  $\varepsilon_{\mu\pi}^{75}$  by about 20% across the momentum bins in comparison with figure 6, which is attributed to  $e^+$  contamination in the training and validation dataset above 40 GeV/ $c$ .

## 7 Summary

A new calorimetric particle identification algorithm based on a convolutional neural network classifier augmented by a filter has been developed for the NA62 experiment at CERN. With respect to the BDT algorithm currently used by NA62, muon misidentification probability as a pion in the momentum range 15–40 GeV/ $c$  is reduced by a factor of six from  $1.2 \times 10^{-5}$  to  $2.0 \times 10^{-6}$ , for a fixed pion-identification efficiency of 75%. Alternatively, pion identification

efficiency is improved in the above momentum range from 72% to 91% for a fixed muon misidentification probability of  $10^{-5}$ . The algorithm is applicable to a wide range of NA62 physics analyses, with best performance achieved after careful parameter tuning.

## Acknowledgments

The authors are grateful to T. Timbers and the University of British Columbia Master of Data Science Capstone team for their assistance. The authors would also like to thank W. Deng, B. Lie and S. Sethi for their contributions to an early phase of this work. It is a pleasure to express our appreciation to the staff of the CERN laboratory and the technical staff of the participating laboratories and universities for their efforts in the operation of the experiment and data processing.

The cost of the experiment and its auxiliary systems was supported by the funding agencies of the Collaboration Institutes. We are particularly indebted to: F.R.S.-FNRS (Fonds de la Recherche Scientifique — FNRS), under Grants No. 4.4512.10, 1.B.258.20, Belgium; CECI (Consortium des Equipements de Calcul Intensif), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region, Belgium; NSERC (Natural Sciences and Engineering Research Council), funding SAPPJ-2018-0017, Canada; MEYS (Ministry of Education, Youth and Sports) funding LM 2018104, Czech Republic; BMBF (Bundesministerium für Bildung und Forschung) contracts 05H12UM5, 05H15UMCNA and 05H18UMCNA, Germany; INFN (Istituto Nazionale di Fisica Nucleare), Italy; MIUR (Ministero dell’Istruzione, dell’Università e della Ricerca), Italy; CONACyT (Consejo Nacional de Ciencia y Tecnología), Mexico; IFA (Institute of Atomic Physics) Romanian CERN-RO No. 1/16.03.2016 and Nucleus Programme PN 19 06 01 04, Romania; MESRS (Ministry of Education, Science, Research and Sport), Slovakia; CERN (European Organization for Nuclear Research), Switzerland; STFC (Science and Technology Facilities Council), United Kingdom; NSF (National Science Foundation) Award Numbers 1506088 and 1806430, U.S.A.; ERC (European Research Council) “UniversaLepto” advanced grant 268062, “KaonLepton” starting grant 336581, Europe.

Individuals have received support from: Charles University Research Center (UNCE/SCI/013), Czech Republic; Ministero dell’Istruzione, dell’Università e della Ricerca (MIUR “Futuro in ricerca 2012” grant RBFR12JF2Z, Project GAP), Italy; the Royal Society (grants UF100308, UF0758946), United Kingdom; STFC (Rutherford fellowships ST/J00412X/1, ST/M005798/1), United Kingdom; ERC (grants 268062, 336581 and starting grant 802836 “AxScale”); EU Horizon 2020 (Marie Skłodowska-Curie grants 701386, 754496, 842407, 893101, 101023808).

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] NA62 collaboration, *Measurement of the very rare  $K^+ \rightarrow \pi^+ \nu \bar{\nu}$  decay*, *JHEP* **06** (2021) 093 [[arXiv:2103.15389](#)] [[INSPIRE](#)].
- [2] A.J. Buras, D. Buttazzo, J. Girrbach-Noe and R. Knegjens,  *$K^+ \rightarrow \pi^+ \nu \bar{\nu}$  and  $K_L \rightarrow \pi^0 \nu \bar{\nu}$  in the Standard Model: status and perspectives*, *JHEP* **11** (2015) 033 [[arXiv:1503.02693](#)] [[INSPIRE](#)].
- [3] P. Baldi, P. Sadowski and D. Whiteson, *Searching for exotic particles in high-energy physics with deep learning*, *Nature Commun.* **5** (2014) 4308 [[arXiv:1402.4735](#)] [[INSPIRE](#)].
- [4] L. de Oliveira et al., *Jet-images — deep learning edition*, *JHEP* **07** (2016) 069 [[arXiv:1511.05190](#)] [[INSPIRE](#)].
- [5] P. Baldi et al., *Jet substructure classification in high-energy physics with deep neural networks*, *Phys. Rev. D* **93** (2016) 094034 [[arXiv:1603.09349](#)] [[INSPIRE](#)].
- [6] G. Ke et al., *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*, in *NIPS* (2017).
- [7] J.H. Friedman, *Greedy function approximation: A gradient boosting machine*, *Annals Statist.* **29** (2001) 1189 [[INSPIRE](#)].
- [8] A. Höecker et al., *TMVA: Toolkit for Multivariate Data Analysis*, *PoS ACAT* (2007) 040 [[physics/0703039](#)].
- [9] NA62 collaboration, *The beam and detector of the NA62 experiment at CERN*, *2017 JINST* **12** P05025 [[arXiv:1703.08501](#)] [[INSPIRE](#)].
- [10] R. Ammendola et al., *The integrated low-level trigger and readout system of the CERN NA62 experiment*, *Nucl. Instrum. Meth. A* **929** (2019) 1 [[arXiv:1903.10200](#)] [[INSPIRE](#)].
- [11] NA62 collaboration, *Performance of the NA62 trigger system*, *JHEP* **03** (2023) 122 [[arXiv:2208.00897](#)] [[INSPIRE](#)].
- [12] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press (2016).
- [13] K.P. Murphy, *Probabilistic Machine Learning: An introduction*, MIT Press (2022).
- [14] K. He, X. Zhang, S. Ren and J. Sun, *Deep residual learning for image recognition*, [arXiv:1512.03385](#) [[DOI:10.1109/CVPR.2016.90](#)] [[INSPIRE](#)].
- [15] NA62 collaboration, *An investigation of the very rare  $K^+ \rightarrow \pi^+ \nu \bar{\nu}$  decay*, *JHEP* **11** (2020) 042 [[arXiv:2007.08218](#)] [[INSPIRE](#)].
- [16] K. He, X. Zhang, S. Ren and J. Sun, *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*, [arXiv:1502.01852](#) [[INSPIRE](#)].
- [17] D.P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, [arXiv:1412.6980](#) [[INSPIRE](#)].
- [18] A. Paszke et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, [arXiv:1912.01703](#).
- [19] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, *Annals Statist.* **7** (1979) 1 [[INSPIRE](#)].

## The NA62 collaboration

### Université Catholique de Louvain, Louvain-La-Neuve, Belgium

E. Cortina Gil , A. Kleimenova <sup>1</sup>, E. Minucci <sup>2</sup>, S. Padolski , P. Petrov, A. Shaikhiev <sup>3</sup>, R. Volpe <sup>4</sup>

### TRIUMF, Vancouver, British Columbia, Canada

W. Fedorko , T. Numao , Y. Petrov , B. Velghe <sup>\*</sup>, V. W. S. Wong , M. Yu 

### University of British Columbia, Vancouver, British Columbia, Canada

D. Bryman <sup>5</sup>, J. Fu

### Charles University, Prague, Czech Republic

Z. Hives , T. Husek , J. Jerhot <sup>6</sup>, K. Kampf , M. Zamkovsky <sup>7</sup>

### Aix Marseille University, CNRS/IN2P3, CPPM, Marseille, France

B. De Martino , M. Perrin-Terrin 

### Institut für Physik and PRISMA Cluster of Excellence, Universität Mainz, Mainz, Germany

A.T. Akmete , R. Aliberti <sup>8</sup>, G. Khorauli <sup>9</sup>, J. Kunze, D. Lomidze <sup>10</sup>, L. Peruzzo , M. Vormstein, R. Wanke 

### Dipartimento di Fisica e Scienze della Terra dell'Università e INFN, Sezione di Ferrara, Ferrara, Italy

P. Dalpiaz, M. Fiorini , A. Mazzolari , I. Neri , A. Norton <sup>11</sup>, F. Petrucci , M. Soldani , H. Wahl <sup>12</sup>

### INFN, Sezione di Ferrara, Ferrara, Italy

L. Bandiera , A. Cotta Ramusino , A. Gianoli , M. Romagnoni , A. Sytov 









### Dipartimento di Fisica e Astronomia dell'Università e INFN, Sezione di Firenze, Sesto Fiorentino, Italy

E. Iacopini , G. Latino , M. Lenti , P. Lo Chiatto , I. Panichi , A. Parenti 

### INFN, Sezione di Firenze, Sesto Fiorentino, Italy

A. Bizzeti <sup>13</sup>, F. Bucci 






### Laboratori Nazionali di Frascati, Frascati, Italy

A. Antonelli , G. Georgiev <sup>14</sup>, V. Kozuharov <sup>14</sup>, G. Lanfranchi , S. Martellotti , M. Moulson , T. Spadaro , G. Tinti 



### Dipartimento di Fisica “Ettore Pancini” e INFN, Sezione di Napoli, Napoli, Italy

F. Ambrosino , T. Capussela, M. Corvino , M. D’Errico , D. Di Filippo , R. Fiorenza <sup>15</sup>, R. Giordano , P. Massarotti , M. Mirra , M. Napolitano , I. Rosa , G. Saracino 










**Dipartimento di Fisica e Geologia dell'Università e INFN, Sezione di Perugia, Perugia, Italy**

G. Anzivino , F. Brizioli <sup>7</sup>, E. Imbergamo, R. Lollini , R. Piandani <sup>16</sup>, C. Santoni 

**INFN, Sezione di Perugia, Perugia, Italy**

M. Barbanera , P. Cenci , B. Checcucci , P. Lubrano , M. Lupi <sup>17</sup>, M. Pepe , M. Piccini 

**Dipartimento di Fisica dell'Università e INFN, Sezione di Pisa, Pisa, Italy**

F. Costantini , L. Di Lella <sup>12</sup>, N. Doble <sup>12</sup>, M. Giorgi , S. Giudici , G. Lamanna , E. Lari , E. Pedreschi , M. Sozzi 

**INFN, Sezione di Pisa, Pisa, Italy**

C. Cerri, R. Fantechi , L. Pontisso <sup>18</sup>, F. Spinella 

**Scuola Normale Superiore e INFN, Sezione di Pisa, Pisa, Italy**

I. Mannelli 

**Dipartimento di Fisica, Sapienza Università di Roma e INFN, Sezione di Roma I, Roma, Italy**

G. D'Agostini , M. Raggi 

**INFN, Sezione di Roma I, Roma, Italy**

A. Biagioni , P. Cretaro , O. Frezza , E. Leonardi , A. Lonardo , M. Turisini , P. Valente , P. Vicini 

**INFN, Sezione di Roma Tor Vergata, Roma, Italy**

R. Ammendola , V. Bonaiuto <sup>19</sup>, A. Fucci, A. Salamon , F. Sargeni <sup>20</sup>

**Dipartimento di Fisica dell'Università e INFN, Sezione di Torino, Torino, Italy**

R. Arcidiacono <sup>21</sup>, B. Bloch-Devauz , M. Boretto <sup>7</sup>, E. Menichetti , E. Migliore , D. Soldi 



**INFN, Sezione di Torino, Torino, Italy**

C. Biino , A. Filippi , F. Marchetto 

**Instituto de Física, Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico**

A. Briano Olvera , J. Engelfried , N. Estrada-Tristan <sup>22</sup>, M. A. Reyes Santos <sup>22</sup>




















**Horia Hulubei National Institute for R&D in Physics and Nuclear Engineering, Bucharest-Magurele, Romania**

P. Boboc , A. M. Bragadireanu, S. A. Ghinescu , O. E. Hutanu


















**Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia**

L. Bician <sup>23</sup>, T. Blazek , V. Cerny , Z. Kucerova <sup>7</sup>



**CERN, European Organization for Nuclear Research, Geneva, Switzerland**

J. Bernhard , A. Ceccucci , M. Ceoletta , H. Danielsson , N. De Simone<sup>24</sup>, F. Duval, B. Döbrich <sup>25</sup>, L. Federici , E. Gamberini , L. Gatignon <sup>3</sup>, R. Guida, F. Hahn<sup>†</sup>, E. B. Holzer , B. Jenninger, M. Koval <sup>23</sup>, P. Laycock <sup>26</sup>, G. Lehmann Miotto , P. Lichard , A. Mapelli , R. Marchevski <sup>1</sup>, K. Massri , M. Noy, V. Palladino , J. Pinzino <sup>27</sup>, V. Ryjov, S. Schuchmann , S. Venditti


**School of Physics and Astronomy, University of Birmingham, Birmingham, United Kingdom**

T. Bache , M. B. Brunetti <sup>28</sup>, V. Duk <sup>4</sup>, V. Fascianelli<sup>29</sup>, J. R. Fry , F. Gonnella , E. Goudzowski , J. Henshaw , L. Iacobuzio, C. Kenworthy , C. Lazzeroni , N. Lurkin <sup>6</sup>, F. Newson, C. Parkinson , A. Romano , J. Sanders , A. Sergi <sup>30</sup>, A. Sturges , J. Swallow <sup>7</sup>, A. Tomczak 

**School of Physics, University of Bristol, Bristol, United Kingdom**

H. Heath , R. Page, S. Trilov 

**School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom**

B. Angelucci, D. Britton , C. Graham , D. Protopopescu 

**Faculty of Science and Technology, University of Lancaster, Lancaster, United Kingdom**

J. Carmignani <sup>31</sup>, J. B. Dainton, R. W. L. Jones , G. Ruggiero <sup>32</sup>


















**School of Physical Sciences, University of Liverpool, Liverpool, United Kingdom**

L. Fulton, D. Hutchcroft , E. Maurice <sup>33</sup>, B. Wrona 

**Physics and Astronomy Department, George Mason University, Fairfax, Virginia, U.S.A.**

A. Conovaloff, P. Cooper, D. Coward <sup>34</sup>, P. Rubin 

**Authors affiliated with an Institute or an international laboratory covered by a cooperation agreement with CERN**

A. Baeva, D. Baigarashev <sup>35</sup>, D. Emelianov, T. Enik , V. Falaleev <sup>4</sup>, S. Fedotov, K. Gorshanov , E. Gushchin , V. Kekelidze , D. Kereibay, S. Kholodenko <sup>27</sup>, A. Khotyantsev, A. Korotkova, Y. Kudenko , V. Kurochka, V. Kurshetsov , L. Litov <sup>14</sup>, D. Madigozhin , M. Medvedeva, A. Mefodev, M. Misheva<sup>36</sup>, N. Molokanova, S. Movchan, V. Obraztsov , A. Okhotnikov , A. Ostankov<sup>†</sup>, I. Polenkevich, Yu. Potrebenikov , A. Sadovskiy , V. Semenov<sup>†</sup>, S. Shkarovskiy, V. Sugonyaev , O. Yushchenko , A. Zinchenko<sup>†</sup>

\* Corresponding author: B. Velghe, email: bvelghe@triumf.ca

<sup>†</sup> Deceased

<sup>1</sup> Present address: Ecole Polytechnique Fédérale Lausanne, CH-1015 Lausanne, Switzerland

<sup>2</sup> Present address: Syracuse University, Syracuse, NY 13244, U.S.A.

- <sup>3</sup> Present address: Faculty of Science and Technology, University of Lancaster, Lancaster, LA1 4YW, U.K.
- <sup>4</sup> Present address: INFN, Sezione di Perugia, I-06100 Perugia, Italy
- <sup>5</sup> Also at TRIUMF, Vancouver, British Columbia, V6T 2A3, Canada
- <sup>6</sup> Present address: Université Catholique de Louvain, B-1348 Louvain-La-Neuve, Belgium
- <sup>7</sup> Present address: CERN, European Organization for Nuclear Research, CH-1211 Geneva 23, Switzerland
- <sup>8</sup> Present address: Institut für Kernphysik and Helmholtz Institute Mainz, Universität Mainz, Mainz, D-55099, Germany
- <sup>9</sup> Present address: Universität Würzburg, D-97070 Würzburg, Germany
- <sup>10</sup> Present address: European XFEL GmbH, D-22869 Schenefeld, Germany
- <sup>11</sup> Present address: School of Physics and Astronomy, University of Glasgow, Glasgow, G12 8QQ, U.K.
- <sup>12</sup> Present address: Institut für Physik and PRISMA Cluster of Excellence, Universität Mainz, D-55099 Mainz, Germany
- <sup>13</sup> Also at Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Università di Modena e Reggio Emilia, I-41125 Modena, Italy
- <sup>14</sup> Also at Faculty of Physics, University of Sofia, BG-1164 Sofia, Bulgaria
- <sup>15</sup> Present address: Scuola Superiore Meridionale e INFN, Sezione di Napoli, I-80138 Napoli, Italy
- <sup>16</sup> Present address: Instituto de Física, Universidad Autónoma de San Luis Potosí, 78240 San Luis Potosí, Mexico
- <sup>17</sup> Present address: Institut am Fachbereich Informatik und Mathematik, Goethe Universität, D-60323 Frankfurt am Main, Germany
- <sup>18</sup> Present address: INFN, Sezione di Roma I, I-00185 Roma, Italy
- <sup>19</sup> Also at Department of Industrial Engineering, University of Roma Tor Vergata, I-00173 Roma, Italy
- <sup>20</sup> Also at Department of Electronic Engineering, University of Roma Tor Vergata, I-00173 Roma, Italy
- <sup>21</sup> Also at Università degli Studi del Piemonte Orientale, I-13100 Vercelli, Italy
- <sup>22</sup> Also at Universidad de Guanajuato, 36000 Guanajuato, Mexico
- <sup>23</sup> Present address: Charles University, 116 36 Prague 1, Czech Republic
- <sup>24</sup> Present address: DESY, D-15738 Zeuthen, Germany
- <sup>25</sup> Present address: Max-Planck-Institut für Physik (Werner-Heisenberg-Institut), München, D-80805, Germany
- <sup>26</sup> Present address: Brookhaven National Laboratory, Upton, NY 11973, U.S.A.
- <sup>27</sup> Present address: INFN, Sezione di Pisa, I-56100 Pisa, Italy
- <sup>28</sup> Present address: Department of Physics, University of Warwick, Coventry, CV4 7AL, U.K.
- <sup>29</sup> Present address: Center for theoretical neuroscience, Columbia University, New York, NY 10027, U.S.A.
- <sup>30</sup> Present address: Dipartimento di Fisica dell'Università e INFN, Sezione di Genova, I-16146 Genova, Italy
- <sup>31</sup> Present address: School of Physical Sciences, University of Liverpool, Liverpool, L69 7ZE, U.K.
- <sup>32</sup> Present address: Dipartimento di Fisica e Astronomia dell'Università e INFN, Sezione di Firenze, I-50019 Sesto Fiorentino, Italy
- <sup>33</sup> Present address: Laboratoire Leprince Ringuet, F-91120 Palaiseau, France
- <sup>34</sup> Also at SLAC National Accelerator Laboratory, Stanford University, Menlo Park, CA 94025, U.S.A.
- <sup>35</sup> Also at L.N. Gumilyov Eurasian National University, 010000 Nur-Sultan, Kazakhstan
- <sup>36</sup> Present address: Institute of Nuclear Research and Nuclear Energy of Bulgarian Academy of Science (INRNE-BAS), BG-1784 Sofia, Bulgaria