COHERENT MUON TO ELECTRON TRANSITION (COMET) PHASE-I LOCAL FILTERING BY CATBOOST ALGORITHM FOR TRACK RECONSTRUCTION

FAHMI IBRAHIM

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Science

Faculty of Science
Universiti Teknologi Malaysia

AUGUST 2022

DEDICATION

This thesis is dedicated to my family. They taught me to be brave and take the risks whatever decision I chosen in my life and facing the reality of life. It is also dedicated to my lab mates and my friends around me, who given me motivation and lesson learn from their previous experiences mainly when I'm facing the problems in my life. They are always offered for help to keep me motivation and finish what I started.

ACKNOWLEDGEMENT

In the first place, I would like to thank Dr. Izyan Hazwani and Dr. Razif Razali and Dr. Nurfikri Norjoharuddeen for their continual support of my physics pursuits and their guidance throughout the years, which has enabled this project to take shape. In doing so, he has given shape to the bright future I now see.

Likewise, I wish to thank Firdaus Soberi for all the hours he spent showing me how to program using C++, especially to make an analysis. I learned countless lessons, best practices, and inspiration from him that I will carry with me for the rest of my life.

To Matthias Dubouchet, Kou Oishi, and Yuki Fujii, thank you for guiding me how to utilize the ICEDUST framework and helping me to overcome technical issues. Your expertise and culture were invaluable to me during the time we spent together. Most of all, thank you for contributing significantly to this work. I couldn't do it without your help.

I wish to thank my Imperial colleagues: Yoshi, Per, Roden and Phill, for all the feedback and wisdom you have given me over the years, and to Yoshi Kuno and the rest of the COMET collaboration, thank you for all the hard work you have put into the project, and thank you for all the fun we had together.

Moving on to institutions, I'd like to thank the National Center for Particle Physics (NCPP), and Data Intensive Computing Centre (DICC) at Universiti Malaya for giving me the opportunity to be Research Assistant at NCPP and providing space to utilize facilities at DICC. Additionally, I'd like to thank cc-IN2P3, France, for allowing to transfer Monte Carlo productions and using the facilities. Furthermore, I would like to thank the Newton-Ungku Omar Fund from the collaboration between Malaysia and the United Kingdom for facilitating my participation in the COMET collaboration, which has allowed me to participate in the physics experiment and follow it up to date.

ABSTRACT

Coherent muon to electron transition (COMET) experiment is an exclusive beamline for studying charge lepton flavor violation through investigation of neutrinoless muon to electron transition. The present work aims to classify signal electron and background from the truth level data generated from GEANT4 simulation (MC5 file) using CatBoost algorithm. This data was first simulated in the Integrated Comet Experimental Data User Software Toolkit (ICEDUST) framework to extract electron and background samples of the main COMET detector, CyDet. Both electron and background samples are merged and the detector response towards this sampling are calibrated using the previous MC4 file. Subsequently, the muon stopping region, bunch width effect, overflow of hits, trigger acceptance and occupancy parameters are observed. The data was sanitized by applying energy cut to the energy deposited on cylindrical drift chamber (CDC) and Cherenkov trigger hodoscope (CTH). Four local features (charge deposited on CDC wire, radial distance of hit from muon stopping target (MST), relative time to the trigger signal, and angle of hit from x-axis) and four neighbour features (charge deposited on right wire, charge deposited on left wire, time relative to the trigger signal on right wire, and time relative to the trigger signal on left wire) are calculated. Using these selected features along with CatBoost algorithm, 94.2% of background hits are removed, whereas 93.7% of hits signal are retained. Performance study using confusion matrix and features importance shows that radial distance from MST gives the highest contribution in the classification of signal and background. Application of machine learning in particle physics is very useful in predicting the experimental sensitivities and processing of big data analysis.

ABSTRAK

Eksperimen peralihan koheren muon ke elektron (COMET) adalah garis alur eksklusif untuk mengkaji pencabulan perisa cas lepton melalui penyiasatan peralihan nyah-neutrino muon ke elektron. Kajian ini bertujuan untuk mengklasifikasi elektron isyarat dan latar belakang daripada data aras kebenaran daripada simulasi GEANT4 (fail MC5) menggunakan algoritma CatBoost. Data ini disimulasikan terdahulu dalam rangka kerja Set Peralatan Perisian Pengguna Data Eksperimen COMET Bersepadu (ICEDUST) untuk mengekstrak sampel elektron dan latar belakang daripada pengesan utama COMET, CyDet. Kedua-dua sampel elektron dan latar belakang digabungkan dan tindak balas pengesan ke atas sampel ini ditentukur menggunakan fail MC4 sebelumnya. Kemudiannya, parameter kawasan perhentian muon, kesan lebar penggugusan, limpahan hits, penerimaan cetusan dan penghunian dicerap. Data ini telah dibersihkan dengan menggunakan potongan tenaga terhadap tenaga termendap di atas kebuk hanyutan silinder (CDC) dan pencetus hodoskop Cherenkov (CTH). Empat ciri-ciri tempatan (cas terkumpul dalam wayar CDC, jarak jejari pukulan dari sasaran muon berhenti (MST), masa relatif terhadap isyarat pencetus, dan sudut pukulan dari paksi-x) dan empat ciri-ciri jiran (cas terkumpul pada wayar kanan, cas terkumpul pada wayar kiri, masa relatif terhadap isyarat pencetus pada wayar kanan, masa relatif terhadap isyarat pencetus pada wayar kiri) dikira. Menggunakan ciri-ciri terpilih ini seiring dengan algoritma CatBoost, 94.2% pukulan latar belakang dialih keluar, manakala 93.7% pukulan isyarat dikekalkan. Kajian prestasi menggunakan matriks kekeliruan dan kepentingan ciri menunjukkan bahawa jarak jejarian dari MST memberikan sumbangan tertinggi dalam klasifikasiisyarat dan latar belakang. Aplikasi pembelajaran mesin dalam zarah fizik sangat berguna dalam meramal sensitiviti eksperimen dan pemprosesan analisa data besar.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADC	-	Analog Digital Converter
AlCap	-	Aluminium Capture
BDT	-	Boosted Decision Tree
CDC	-	Cylindrical Drift Chamber
CDCHF	-	Cylindrical Drift Chamber Hit Filter
CFRP	-	Carbon Fiber Reinforced Plastic
CLFV	-	Charge Lepton Flavor Violation
CPU	-	Central Processing Unit
COMET	-	Coherent Muon-to-Electron Transition
CTH	-	Cherenkov Trigger Hodoscope
CyDet	-	Cylindrical Detector
CV	-	Cross Validation
DA	-	Data Analiytics
DAQ	-	Digital Analog Converter
DIO	-	Decay In Orbit
DNN	-	Deep Neural Network
ECAL	-	Electron Calorimeter
GBDT	-	Gradient Boosted Decision Tree
GPU	-	Graphical Processing Unit
HPC	-	High Performance Computing
ICEDUST	-	Integrated COMET Experimental Data User Software
	-	Toolkit
IFN	-	Information Fuzzy Network
J-PARC	-	Japan Proton Accelerator Research Complex
KEK	-	Kō Enerugī Kasokuki Kenkyū Kikō
LHC	-	Large Hadron Collider

MC	-	Monte Carlo
ML	-	Machine Learning
MR	-	Main Ring
MST	-	Muon Stopping Target
ND	-	Near Detector
NLFV	-	Neutral Lepton Flavor Violation
NP	-	Nuclear Physics
PID	-	Particle Identification
PMT	-	Photomultiplier
PP	-	Pion Production
POT	-	Proton On Target
PSI	-	Paul Schrerrer Institutte
PSM	-	Physics Standard Model
ROC	-	Receiver Operating Characteristics
ROC-AUC	-	Receiver Operating Characteristics - Area Under Curve
SES	-	Single Event Sensitivity
SM	-	Standard Model
STL	-	Standard Template Library
StrEcal	-	Straw Tracker And The Electron Calorimeter
TB	-	Terabyte

LIST OF SYMBOLS

A	-	Atomic Number
Al	-	Aluminium
Au	-	Gold
B	-	Magnetic Field
B_{μ}	-	Muon Binding Energy
B_I	-	Bunch Intensity
BN	-	Bunch Number
B_T	-	Time Bunch Separation
B_w	-	Bunch Width
c	-	Speed of Light
$E_{\mu e}$	-	Energy of Electron Signal
E_{recoil}	-	Recoil Energy
E_{dep}	-	Energy Deposit
K	-	Kaon
L	-	Lepton Number
L_e	-	Lepton Number of Electron
L_{μ}	-	Lepton Number of Muon
L_{τ}	-	Lepton Number of Taon
n	-	Refractive Index
n	-	neutron
p	-	proton
N	-	Nucleon Number
m_{μ}	-	Rest Mass of Muon
$p_{\perp}$	-	Perpendicular Transverse Momentum
$p_{\parallel}$	-	Parallel Transverse Momentum
Pb	-	Lead

q	-	Charge Deposit
q_R	-	Charge Deposit on Right Wire
q_L	-	Charge Deposit on Left Wire
R	-	Radius of Curvature
t_{RO}	-	The Time of Readout Recorded
t_D	-	The Drift Time
t_{DAQ}	-	Time of DAQ
t_H	-	The Time of Hit Recorded
t_R	-	Time of Relative Hit To The Trigger Signal
t_{RR}	-	Time of Relative Hit To The Trigger Signal on Right Wire
t_{RL}	-	Time of Relative Hit To The Trigger Signal on Left Wire
v	-	Velocity of Particle
y_{local}	-	Predicted Output of Local CatBoost
$y_{neighbor}$	-	Predicted Output of Neighbor CatBoost
Z	-	Proton Number
γ	-	Gamma
e^-	-	Electron
μ^-	-	Muon
ϕ	-	Angle of Hit from x-axis
$\|x\|$	-	Radial Distance of Hit from MST
β	-	Lorentz Factor

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of Study

The standard model (SM) of physics includes six flavors of quarks, six leptons, and four interaction forces. However, the discovery of the Higgs boson at the Large Hadron Collider (LHC) and neutrino mass observation demonstrates the clear evidence that some modification is necessary for the SM [1]. Further study may unleash the answer for unresolved questions in particle physics such as domination of matter in the universe, the existence of dark matter, quantum gravity problem and neutrino mass [2, 3, 4]. The present Coherent Muon to Electron Transition (COMET) experiment aims to study the charged lepton flavor violation by observing the neutrinoless muon to electron conversion.

The common allowed process of nuclear muon capture is when a negative muon being capture in a target $\mu^- + N(A, Z) \rightarrow e^- + N(A, Z - 1)$ and formed a muonic atom. The muon is bound at certain energy levels before cascades down to its $1s$ ground state. The normal expected event for muon decays into electron within an average of 2.2×10^{-6} s. However, the neutrino with tendencies to oscillate and unknown symmetry of the neutrino causes the rare neutrinoless muon to electron conversion. The coherent neutrinoless conversion of muon to electron decay requires an explanation of physics beyond SM since it violates the conservation of individual lepton flavors L_e and L_μ by 1 unit. However, the total lepton number $L = L_e + L_\mu + L_\tau$ of the reaction is conserved.

The signal electron energy, $E_{\mu e}$ is given by:

$$E_{\mu e} = m_\mu - B_\mu - E_{recoil} \tag{1.1}$$

where m_μ is the mass of the muon, B_μ is the binding energy of the $1s$-state muonic atom, and E_{recoil} is the nuclear recoil energy. The $E_{\mu e}$ is dependent on the material. For example, for the COMET experiment that uses Al as muon stopping target, the $E_{\mu e}$ is 104.97 MeV. Alternatively, the $E_{\mu e}$ of lead (Pb) is 94.9 MeV.

The COMET experiment uses two-stage approaches to measure the $\mu - e$ conversion with an unprecedented sensitivity of 10^{-15} to 10^{-17} in COMET Phase-I and Phase-II, respectively. This ultimate sensitivity goal is a factor of about 10^4 better than the current experimental limit of $B(\mu^- + Au \rightarrow e^- + Au) \leq 7\times10^{-13}$) from SINDRUM-II at Paul Scherrer Institute (PSI) [5]. A huge number of muons are required in order to achieve this sensitivity. At this moment, the COMET Phase-I is under construction at Japan Particle Accelerator Research Complex (J-PARC), Tokai, and the preparation of the experiment is intensively in progress [6]. The construction of the COMET building is equipped with a 90° bending angle of transport solenoid installed in the COMET hall. The central drift chamber (CDC) is the main detector in COMET Phase-I is separately constructed and test using cosmic rays at the High Energy Accelerator Research Organization (KEK), Tsukuba. The calibration and tracking efficiency of the CDC is done using various machine learning to check the reproducibility of the signal tracking and data reconstruction.

Various boosting algorithms are widely used in the application of particle physics. The gradient boosted decision trees (GBDT) [7] is one of the well-known machine learning algorithms that give high-quality models in a huge number of machine learning problems involving heterogeneous features, noisy data, and complex dependencies [8, 9]. For instance, search engines [10, 11], recommendation systems [12], and other applications [13, 14] uses GBDT algorithm to predict users need. CatBoost is one of the third-party GBDTs algorithms that implement gradient boosting by using oblivious decision trees as base predictors[15]. LHCb experiment uses CatBoost algorithm to classify the particle identification (PID) for their detectors. The preliminary results from 60 observable features in LHCb systems compare between CatBoost, deep neural network (DNN), Flat 4d and baseline (ProbNN) demonstrates that CatBoost and DNN models have shown good performance compare to another algorithm. In other work by [16], CatBoost and DNN also outperform the ProbNN model for the study of p-vs-K pair.

1.2 Problem Statement

The main issue for this high sensitivities detection at COMET is that the truth level data from simulation should have very high statistics. The COMET collaboration updated their Monte Carlo (MC) simulation file. The gap year to move from MC4 to MC5 is 2 years and there are a lot of changes in MC5 especially geometry updates and the method of the simulation. The present MC5 file uses combinations of pre-built physics processes registered in Geant4 and additional custom physics models based on the latest experimental data. The generated file described all relevant physics processes and thus produced many other particles mark with their ID number. Furthermore, the low energy physics involving neutrons and other hadronic scattering given by the QGSP_BERT_HP program is the best accuracy for the COMET experiment to improve low-energy electromagnetic interactions [17]. The major source of background in COMET Phase-I is muon nuclear capture in aluminium. The AlCap experiment have been discussed about the resulting spectra from this process [18]. The charged particle emission from nuclear muon capture on aluminium at high energies (E > 40 MeV) [19]. This is too high for COMET. Whereas at low energies is only to know for muonic silicon [20]. The rate of neutrons is known [21], but the spectrum is undefined. The rates and energies of photons, X-ray and gamma ray are known. However, there are huge uncertainties on the gamma ray intensities [22]. The proton emission frequency and spectrum from AlCap experiment implemented into the custom stopped muon physics model. The Geant4 uses a Bertini cascade to model this process and resulting roughly 20% of proton emission of all captured muons. Whilst the AlCap predicted a much lower rate of proton emission about 3%. Obviously, the spectrum measured by Bertini cascade is more energetic than the spectrum measured by AlCap. Hence, MC5 has higher the total number proton on target (POT) is higher than MC4. This can be summarized in the Table 1.1 :

Secondly, the scheduled event for neutrinoless muon to electron conversion is quite high since the used muon intensity is about 10^{11} muons/s that is about 10^5

Table 1.1: Summary of MC4o and MC5.

	MC4o	MC5
Simulation	Simulating particles from the Pion Production Target into the Muon Transport or Detector Solenoid sections is recorded so that it can be resampled later.	Simulation is split into two parts a) upstream (Pion Production Target) and b) downstream (CyDet).
Resampling	Yes	No
Phase-I Sampling World	No	Yes
Physic list	QGSP_BERT_HP	
Total proton on target (POT)	500,000,000	990,677,775
POT per bunch	16,000,000	8,000,000

order higher than the simulated events. Machine learning encompasses automatic computing procedures based on logical or binary operations that learn a task from a series of examples through a process of interference and model fitting [25, 26]. In this study, supervised machine learning is used on MC5's electron signal to classify the backgrounds and electron signal. The CatBoost is a third-party package of GBDTs algorithms like XGBoost and LightGBM that reported up-to-date comparison of state-of-art classification algorithms by Zhang et al. [27]. The implementation of oblivious decision trees in Catboost makes them effective for feature selection despite other restrictions applied in the CatBoost. The factual difference between the oblivious decision tree and a regular decision tree is vital to minimize the overall subset of input attributes. The previous work of the CDC track reconstruction consists of three main stages: a) local filtering, b) shape recognition and c) track filtering. In this research, the track reconstruction is only focused on the local filtering by local and neighbouring classification tasks. The local filtering is used as many features as possible to build a classifier in order to select signal hits.

1.3 Objectives of Study

The objectives of the research are as follows:

i. To determine estimate trigger acceptance of signal and background and their occupancy.

ii. To determine local and neighbor features based on local detector information from ICEDUST simulation.

iii. To evaluate the overall performance of the neutrinoless muon to electron decay signal with background using CatBoost algorithm.

1.4 Scope of Study

The present work uses the latest MC5 file from the monte Carlo simulation exclusive for the COMET experiment available in the IN2P3 cluster. The current MC5 file consists of two parts which are MC5A01 for upstream (simulation before interaction with the muon stopping target (MST)) and MC5A02 for downstream (simulation upon interaction with MST and beyond). Further details related to the MC5 file will be discussed in the section 2.3.9. The current study focuses on research in the CyDet detector region of COMET Phase-I. Thus MC5A02 file is used for the present analysis.

The ICEDUST simulation is started with GEANT4 simulation through SimG4 by giving MC5A02 as input in this simulation. In the SimHitMerger phase, the number of POT per bunch is merged to 8×10^6 and the bunch separation between bunches is set to 1170 ns based on COMET TDR. Thus the number of muons stopped in the Al disk per POT event is calculated and the effect of bunch width of proton beam, B_w on stopped muons, CTH and CDC timings. The output of SimHitMerger is feed to SimDetectorResponse. At this phase, the hits in the CTH and CDC detectors are calibrated like real measurements. The outputs from this simulation are studied such as drift time of CDC, the overflow rates in CTH and CDC and their causes, the trigger of acceptance signal and background and their occupancies are determined.

The pre-processing data is significant. The signal and background are defined. Afterward, the data extraction is performed to extract data from detector response simulation and data sanitization is required to remove irrelevant information. The optimization of CTH and CDC samples are optimized according to detector information such as energy deposited in the detector and the timing of trigger window. The feature selection and calculation are performed. These create the 4 of local features and the other 4 of neighbor features.

Since this is supervised machine learning, the background and signal samples are prepared by providing local and neighbor features. The samples are split into 60% of training set and 40% testing set. The CatBoost classifier is used and the hyperparameters are optimized such as learning rate, depth of tree, number of trees and type of loss function. The classification between electron signals and backgrounds is done using supervised machine learning through the CatBoost algorithm. The performances of CatBoost and features evaluation are evaluated through classification outputs, confusion matrix, ROC and feature importance. From the ROC, the optimization the trade-off between background rejection efficiency and signal retention efficiency are determined using G-mean. The optimal background rejection efficiency and signal retention efficiency are determined.

1.5 Significance of Study

The first physics run of COMET Phase-I is expected in March 2023. Estimation by simulation is necessary before the real experiment is carried out to achieve the desired single event sensitivity (SES). The success of the COMET experiment will open new opportunities to physics beyond SM, where a positive result for this measurement would be a Nobel-prize level discovery. While the absence of the CLFV signal is also a high-impact result, COMET is improving on the current sensitivity by a factor of 10,000 from the SINDRUM-II experiment at PSI, Switzerland. Furthermore, the observation from charged lepton flavor violation through this neutrinoless muon to electron decay can confirm the neutrino mixing postulates in Kobayashi-Maskawa's prediction. The construction of the COMET facility in J-PARC could become the hub of high energy physics for nuclear and particle physics and other related fields.

Machine-learning approaches for data analysis in particle physics experiment has been widely used due to expensive computational cost. However, big data analysis is needed to predict particle classification by giving target variables either "0"(background) or "1"(signal) to save more time and train larger datasets. Malaysia is currently in the midst of its own Data Analytics (DA) revolution. Through governmental policy, it supports both academic research and industrial avenues to increase Malaysia's capacity for DA and High-Performance Computing (HPC).

1.6 Thesis Outline

This research report is organized as following. The background study about the MC5 and CatBoost algorithm are proposed for tracking reconstruction of muon to electron conversion. The literature review is focusing on COMET Phase-I and its current status, the official software for COMET experiment which is called ICEDUST, the detail about MC5 productions, introduce the fundamental of machine learning and Gradient Boosted Decision Trees algorithm (GBDT), which is CatBoost algorithm, and the overview of previous work using MC4o production in order to reconstruct the track muon to electron conversion using local and neighbouring features. Next, the methodology about ICEDUST simulation from initial proton on target through SimG4, followed by merging the event into bunch-like form in the SimHitMerger, and the simulation of detector response via SimDetectorResponse to produce mimic like experimental data. Then the implementation of CatBoost algorithm for the tracking study for classification between signal hits and background hits using local and neighbouring features. Furthermore, the results of simulation are discussed about the muons stopping target distribution, bunch width effect, detector hits timing, analysis of timing, triggering and tracking cuts, and the performance of CatBoost algorithm is evaluated for tracking reconstruction of muon to electron conversion for COMET Phase-I. The conclusion have been summarize the thesis, restate the contributions of this model, limitations and recommendation in order to improve this model.

REFERENCES

1. Kuno, Y. A search for muon-to-electron conversion at J-PARC: the COMET experiment. *Progress of Theoretical and Experimental Physics*, 2013. 2013(2): 022C01.

2. Bertone, G., Hooper, D. and Silk, J. Particle dark matter: evidence, candidates and constraints. *Physics reports*, 2005. 405(5-6): 279–390.

3. Fukuda, Y., Hayakawa, T., Ichihara, E., Inoue, K., Ishihara, K., Ishino, H., Itow, Y., Kajita, T., Kameda, J., Kasuga, S. *et al.* Evidence for oscillation of atmospheric neutrinos. *Physical Review Letters*, 1998. 81(8): 1562.

4. Fukuda, S., Fukuda, Y., Ishitsuka, M., Itow, Y., Kajita, T., Kameda, J., Kaneyuki, K., Kobayashi, K., Koshio, Y., Miura, M. *et al.* Constraints on neutrino oscillations using 1258 days of Super-Kamiokande solar neutrino data. *Physical Review Letters*, 2001. 86(25): 5656.

5. Bertl, W., Engfer, R., Hermes, E., Kurz, G., Kozlowski, T., Kuth, J., Otter, G., Rosenbaum, F., Ryskulov, N., Van Der Schaaf, A. *et al.* A search for μ-e conversion in muonic gold. *The European Physical Journal C-Particles and Fields*, 2006. 47(2): 337–346.

6. Litchfield, R. Status of the COMET experiment. *EPJ Web of Conferences*. EDP Sciences. 2018, vol. 179. 01011.

7. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 2001: 1189–1232.

8. Roe, B. P., Yang, H.-J., Zhu, J., Liu, Y., Stancu, I. and McGregor, G. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2005. 543(2-3): 577–584.

9. Ibragimov, B. and Gusev, G. Minimal variance sampling in stochastic gradient boosting. *arXiv preprint arXiv:1910.13204*, 2019.

10. Wu, Q., Burges, C. J., Svore, K. M. and Gao, J. Adapting boosting for information retrieval measures. *Information Retrieval*, 2010. 13(3): 254–270.

11. Burges, C. J. From ranknet to lambdarank to lambdamart: An overview. *Learning*, 2010. 11(23-581): 81.

12. Richardson, M., Dominowska, E. and Ragno, R. Predicting clicks: estimating the click-through rate for new ads. *Proceedings of the 16th international conference on World Wide Web*. 2007. 521–530.

13. Zhang, Y. and Haghani, A. A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, 2015. 58: 308–324.

14. Caruana, R. and Niculescu-Mizil, A. An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on Machine learning*. 2006. 161–168.

15. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. and Gulin, A. CatBoost: unbiased boosting with categorical features. *arXiv preprint arXiv:1706.09516*, 2017.

16. Derkach, D., Hushchyn, M. and Kazeev, N. Machine Learning based Global Particle Identification Algorithms at the LHCb Experiment. *EPJ Web of Conferences*. EDP Sciences. 2019, vol. 214. 06011.

17. Edmonds, A. W. J. *An Estimate of the Hadron Production Uncertainty and a Measurement of the Rate of Proton Emission after Nuclear Muon Capture for the Comet Experiment*. Ph.D. Thesis. UCL (University College London). 2015.

18. Edmonds, A. Latest Updates from the AlCap Experiment. *arXiv preprint arXiv:1809.10122*, 2018.

19. Wyttenbach, A., Baertschi, P., Bajo, S., Hadermann, J., Junker, K., Katcoff, S., Hermes, E. and Pruys, H. Probabilities of muon induced nuclear reactions involving charged particle emission. *Nuclear Physics A*, 1978. 294(3): 278–292.

20. Sobottka, S. E. and Wills, E. L. Energy Spectrum of Charged Particles Emitted Following Muon Capture in Si 28. *Physical Review Letters*, 1968. 20(12): 596.

21. Macdonald, B., Diaz, J. A., Kaplan, S. N. and Pyle, R. V. Neutrons from negative-muon capture. *Physical Review*, 1965. 139(5B): B1253.

22. Measday, D. F., Stocki, T. J., Moftah, B. A. and Tam, H. γ rays from muon capture in Al 27 and natural Si. *Physical Review C*, 2007. 76(3): 035504.

23. Sam, W. MC4o Production, 2017. URL `https://https://gitlab.in2p3.fr/comet/ICEDUST_productions/-/wikis/Mc4o`.

24. Doubchet, M. MC5 Production, 2020. URL `https://gitlab.in2p3.fr/comet/ICEDUST_productions/-/wikis/MC5A`.

25. Michie, D., Spiegelhalter, D. J. and Taylor, C. C. Machine learning, neural and statistical classification. 1994.

26. Ayodele, T. O. Machine learning overview. *New Advances in Machine Learning*, 2010: 9–19.

27. Zhang, C., Liu, C., Zhang, X. and Almpanidis, G. An up-to-date comparison of state-of-the-art classification algorithms. *Expert Systems with Applications*, 2017. 82: 128–150.

28. Nishina, T. Ichimiya 1937 Phys. *Rev*. 52: 1198.

29. Danby, G., Gaillard, J. M., Goulianos, K., Lederman, L. M., Mistry, N., Schwartz, M. and Steinberger, J. Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos. *Physical Review Letters*, 1962. 9(1): 36.

30. Krommes, J. A. The gyrokinetic description of microturbulence in magnetized plasmas. *Annual review of fluid mechanics*, 2012. 44: 175–201.

31. Davis Jr, R., Harmer, D. S. and Hoffman, K. C. Search for neutrinos from the sun. *Physical Review Letters*, 1968. 20(21): 1205.

32. Pontecorvo, B. Neutrino experiments and the problem of conservation of leptonic charge. *Sov. Phys. JETP*, 1968. 26(984-988): 165.

33. Collaboration, S.-K. *et al.* Y. Fukuda et al., Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett*, 1998. 81(8): 1562–1567.

34. Hincks, E. and Pontecorvo, B. Search for gamma-radiation in the 2.2-microsecond meson decay process. *Physical Review*, 1948. 73(3): 257.

35. Bernstein, R. H. and Cooper, P. S. Charged lepton flavor violation: an experimenter's guide. *Physics Reports*, 2013. 532(2): 27–64.

36. collaboration, M. *et al.* Search for the Lepton Flavour Violating Decay $\mu^+ \rightarrow e^+\gamma$ with the Full Dataset of the MEG Experiment. *arXiv preprint arXiv:1605.05081*, 2016.

37. Hincks, E. and Pontecorvo, B. On the absence of photons among the decay products of the 2.2 microsecond meson. *Canadian Journal of Research*, 1950. 28(1): 29–43.

38. Adam, J., Bai, X., Baldini, A. M., Baracchini, E., Bemporad, C., Boca, G., Cattaneo, P. W., Cavoto, G., Cei, F., Cerri, C. *et al.* New constraint on the existence of the μ+→ e+ γ decay. *Physical Review Letters*, 2013. 110(20): 201801.

39. Adam, J., Bai, X., Baldini, A., Baracchini, E., Bemporad, C., Boca, G., Cattaneo, P., Cavoto, G., Cei, F., Cerri, C. *et al.* New limit on the lepton-flavor-violating decay μ+→ e+ γ. *Physical review letters*, 2011. 107(17): 171801.

40. Blondel, A., Bravar, A., Pohl, M., Bachmann, S., Berger, N., Kiehn, M., Schöning, A., Wiedner, D., Windelband, B., Eckert, P. *et al.* Research Proposal for an Experiment to Search for the Decay {\mu}-> eee. *arXiv preprint arXiv:1301.6113*, 2013.

41. Bellgardt, U., Otter, G., Eichler, R., Felawka, L., Niebuhr, C., Walter, H., Bertl, W., Lordong, N., Martino, J., Egli, S. *et al.* Search for the decay μ+→ e+ e+ e-. *Nuclear Physics B*, 1988. 299(1): 1–6.

42. Kaulard, J., Dohmen, C., Haan, H., Honecker, W., Junker, D., Otter, G., Starlinger, M., Wintz, P., Hofmann, J., Bertl, W. *et al.* Improved limit on the branching ratio of μ-→ e+ conversion on titanium. *Physics Letters B*, 1998. 422(1-4): 334–338.

43. Bertl, W., Engfer, R., Hermes, E., Kurz, G., Kozlowski, T., Kuth, J., Otter, G., Rosenbaum, F., Ryskulov, N., Van Der Schaaf, A. *et al.* A search for μ-e conversion in muonic gold. *The European Physical Journal C-Particles and Fields*, 2006. 47(2): 337–346.

44. Bryman, D., Blecher, M., Gotow, K. and Powers, R. Search for the Reaction μ-+ Cu→ e++ Co. *Physical Review Letters*, 1972. 28(22): 1469.

45. Badertscher, A., Borer, K., Czapek, G., Flückiger, A., Hänni, H., Hahn, B., Hugentobler, E., Markees, A., Marti, T., Moser, U. *et al.* A search for muon-electron and muon-positron conversion in sulfur. *Nuclear Physics A*, 1982. 377(2-3): 406–440.

46. Bryman, D., Clifford, E., Leitch, M., Navon, I., Numao, T., Schlatter, P., Dixit, M., Hargrove, C., Mes, H., Burnham, R. *et al.* Search for μ-e conversion in Ti. *Physical Review Letters*, 1985. 55(5): 465.

47. Ahmad, S., Azuelos, G., Blecher, M., Bryman, D., Burnham, R., Clifford, E., Depommier, P., Dixit, M., Gotow, K., Hargrove, C. *et al.* Search for muon-electron and muon-positron conversion. *Physical Review D*, 1988. 38(7): 2102.

48. Dohmen, C., Groth, K.-D., Heer, B., Honecker, W., Otter, G., Steinrücken, B., Wintz, P., Djordjadze, V., Hofmann, J., Kozlowski, T. *et al.* Test of lepton-flavour conservation in $\mu\to$ e conversion on titanium. *Physics Letters B*, 1993. 317(4): 631–636.

49. Honecker, W., Dohmen, C., Haan, H., Junker, D., Otter, G., Starlinger, M., Wintz, P., Hofmann, J., Bertl, W., Egger, J. *et al.* Improved limit on the branching ratio of $\mu\to$ e conversion on lead. *Physical review letters*, 1996. 76(2): 200.

50. Wintz, P. Proceedings of the First International Symposium on Lepton and Baryon Number Violation. 1998.

51. Lee, M. COMET muon conversion experiment in J-PARC. *Frontiers in Physics*, 2018. 6: 133.

52. Teshima, N. μ-e conversion experiments at J-PARC. *Proceedings for the 15th International Workshop on Tau Lepton Physics*. 2018, vol. 24. 28.

53. Tran, N. H. A study of proton emission following nuclear muon capture for the COMET experiment. 2014.

54. Zuber, K., Zhang, Y., Zhang, J., Zdorovets, M., Yudin, Y. V., Yuan, Y., Yoshioka, T., Yoshida, M., Yoshida, H., Yeo, B. *et al.* COMET Phase-I technical design report. *Progress of Theoretical and Experimental Physics*, 2020. 2020(3).

55. Measday, D. F. The nuclear physics of muon capture. *Physics Reports*, 2001. 354(4-5): 243–409.

56. Zinatulina, D., Brudanin, V., Egorov, V., Petitjean, C., Shirchenko, M., Suhonen, J. and Yutlandov, I. Ordinary muon capture studies for the matrix elements in $\beta\beta$ decay. *Physical Review C*, 2019. 99(2): 024327.

57. Šimkovic, F., Dvornickỳ, R. and Vogel, P. Muon capture rates: Evaluation within the quasiparticle random phase approximation. *Physical Review C*, 2020. 102(3): 034301.

58. Dietz-Laursonn, E. Peculiarities in the simulation of optical physics with GEANT4. *arXiv preprint arXiv:1612.05162*, 2016.

59. Czarnecki, A., i Tormo, X. G. and Marciano, W. J. Muon decay in orbit: spectrum of high-energy electrons. *Physical Review D*, 2011. 84(1): 013006.

60. Böhlen, T., Cerutti, F., Chin, M., Fassò, A., Ferrari, A., Ortega, P. G., Mairani, A., Sala, P. R., Smirnov, G. and Vlachoudis, V. The FLUKA code: developments and challenges for high energy and medical applications. *Nuclear data sheets*, 2014. 120: 211–214.

61. Iwase, H., Niita, K. and Nakamura, T. Development of general-purpose particle and heavy ion transport Monte Carlo code. *Journal of Nuclear Science and Technology*, 2002. 39(11): 1142–1151.

62. Kurup, A., Puri, I., Uchida, Y., Yap, Y., Appleby, R., Tygier, S., D'Arcy, R., Edmonds, A., Lancaster, M. and Wing, M. Large Emittance Beam Measurements for COMET Phase-I. *Proceedings, 4th International Particle Accelerator Conference (IPAC 2013).* 2013. 2684.

63. Höppner, C., Neubert, S., Ketzer, B. and Paul, S. A novel generic framework for track fitting in complex detector systems. *Nuclear Instruments*

and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2010. 620(2-3): 518–525.

64. Radovic, A., Williams, M., Rousseau, D., Kagan, M., Bonacorsi, D., Himmel, A., Aurisano, A., Terao, K. and Wongjirad, T. Machine learning at the energy and intensity frontiers of particle physics. *Nature*, 2018. 560(7716): 41–48.

65. Breiman, L., Friedman, J., Stone, C. J. and Olshen, R. A. *Classification and regression trees*. CRC press. 1984.

66. Freund, Y. and Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 1997. 55(1): 119–139.

67. Lönnblad, L., Peterson, C. and Rögnvaldsson, T. Finding gluon jets with a neural trigger. *Physical review letters*, 1990. 65(11): 1321.

68. Babbage, W. S. and Thompson, L. F. The use of neural networks in 0 discrimination. *Nucl. Instrum. Methods A*, 1993. 330: 482.

69. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. and Gulin, A. CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems*. 2018. 6638–6648.

70. Hancock, J. T. and Khoshgoftaar, T. M. CatBoost for big data: an interdisciplinary review. *Journal of big Data*, 2020. 7(1): 1–45.

71. Rokach, L. and Maimon, O. Top-down induction of decision trees classifiers-a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 2005. 35(4): 476–487.

72. Almuallim, H. and Dietterich, T. G. Learning boolean concepts in the presence of many irrelevant features. *Artificial intelligence*, 1994. 69(1-2): 279–305.

73. Schlimmer, J. C. *et al.* Efficiently inducing determinations: A complete and systematic search algorithm that uses optimal pruning. *Proceedings of the 1993 International Conference on Machine Learning*. 1993. 284–290.

74. Langley, P. and Sage, S. Oblivious decision trees and abstract cases. *Working notes of the AAAI-94 workshop on case-based reasoning*. Seattle, WA. 1994. 113–117.

75. Kohavi, R. and Sommerfield, D. Targeting Business Users with Decision Table Classifiers. *KDD*. 1998. 249–253.

76. Last, M., Maimon, O. and Minkov, E. Improving stability of decision trees. *International Journal of Pattern Recognition and Artificial Intelligence*, 2002. 16(02): 145–159.

77. Psihas, F., Groh, M., Tunnell, C. and Warburton, K. A review on machine learning for neutrino experiments. *International Journal of Modern Physics A*, 2020. 35(33): 2043005.

78. Brun, R. and Rademakers, F. ROOT—An object oriented data analysis framework. *Nuclear instruments and methods in physics research section A: accelerators, spectrometers, detectors and associated equipment*, 1997. 389(1-2): 81–86.

79. Chen, T. and Guestrin, C. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016. 785–794.

80. Al Daoud, E. Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *International Journal of Computer and Information Engineering*, 2019. 13(1): 6–10.

81. Gillies, E. COMET Phase-I track reconstruction using machine learning and computer vision. 2018.

82. Jones, E., Oliphant, T., Peterson, P. *et al.* SciPy: Open source scientific tools for Python. 2001.

83. McKinney, W. *et al.* Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*. Austin, TX. 2010, vol. 445. 51–56.

84. Chen, T. *et al.* guestrin, C. Xgboost: A scalable tree boosting system. *Proceedings of the 22Nd ACM SIGKDD International Conference on*

Knowledge Discovery and Data Mining (New York, NY, USA, 2016), KDD '16, ACM. 2016. 785–794.

85. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 2017. 30.

86. Murphy, C. W. Class imbalance techniques for high energy physics. *SciPost Physics*, 2019. 7(6): 076.

87. Bentéjac, C., Csörgő, A. and Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 2021. 54(3): 1937–1967.

88. Lindner, T. Evolution of the T2K-ND280 Computing Model. *Journal of Physics: Conference Series*. IOP Publishing. 2015, vol. 664. 032021.

89. Brown, K. L. and Iselin, F. C. *DECAY TURTLE (Trace Unlimted Rays Through Lumped Elements): a computer program for simulating charged-particle beam transport systems, including decay calculations*. Technical report. CERN. 1974.

90. Collaboration, G., Agostinelli, S. *et al.* GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 2003. 506(25): 0.

91. Marom, N. D., Rokach, L. and Shmilovici, A. Using the confusion matrix for improving ensemble classifiers. *2010 IEEE 26-th Convention of Electrical and Electronics Engineers in Israel*. IEEE. 2010. 000555–000559.

92. McClish, D. K. Analyzing a portion of the ROC curve. *Medical decision making*, 1989. 9(3): 190–195.

93. Flach, P. A. ROC analysis. In: *Encyclopedia of Machine Learning and Data Mining*. Springer. 1–8. 2016.

94. Hoo, Z. H., Candlish, J. and Teare, D. What is an ROC curve?, 2017.

95. Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J. and Kattan, M. W. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.)*, 2010. 21(1): 128.

96. Xu, J., Zhang, Y. and Miao, D. Three-way confusion matrix for classification: A measure driven view. *Information sciences*, 2020. 507: 772–794.

97. Maria Navin, J. and Pankaja, R. Performance analysis of text classification algorithms using confusion matrix. *International Journal of Engineering and Technical Research (IJETR)*, 2016. 6(4): 75–8.

98. Kim, M.-J., Kang, D.-K. and Kim, H. B. Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. *Expert Systems with Applications*, 2015. 42(3): 1074–1082.

99. Kubat, M., Holte, R. and Matwin, S. Learning when negative examples abound. *European conference on machine learning*. Springer. 1997. 146–153.

100. Fawcett, T. An introduction to ROC analysis. *Pattern recognition letters*, 2006. 27(8): 861–874.

101. Kang, P., Lin, Z., Teng, S., Zhang, G., Guo, L. and Zhang, W. Catboost-based framework with additional user information for social media popularity prediction. *Proceedings of the 27th ACM international conference on multimedia*. 2019. 2677–2681.

102. Bernstein, R. H. The mu2e experiment. *Frontiers in Physics*, 2019. 7: 1.

103. Suzuki, T., Measday, D. F. and Roalsvig, J. Total nuclear capture rates for negative muons. *Physical Review C*, 1987. 35(6): 2212.

104. Frü'hwirth, R., Regler, M., Bock, R., Grote, H. and Notz, D. *Data analysis techniques for high-energy physics*. 2000.

105. Andrews, H. C. *Introduction to mathematical techniques in pattern recognition*. RE Krieger Pub. Co. 1983.

106. Edmonds, A., Brown, D., Vinas, L. and Pagan, S. Using machine learning to select high-quality measurements. *Journal of Instrumentation*, 2021. 16(08): T08010.

107. Saarimäkio, O. USING BOOSTED DECISION TREES IN THE COMET EXPERIMENT, 2016.

108. Bock, R. K., Grote, H. and Notz, D. *Data analysis techniques for high-energy physics*. vol. 11. Cambridge University Press. 2000.

109. Schmidt, B. The High-Luminosity upgrade of the LHC: Physics and Technology Challenges for the Accelerator and the Experiments. *Journal of Physics: Conference Series*. IOP Publishing. 2016, vol. 706. 022002.

LIST OF PUBLICATIONS

Journal with Impact Factor

1. I.H. Hashim, H. Ejiri, F. Othman, **F. Ibrahim**, F. Soberi, N.N.A.M.A. Ghani, T. Shima, A. Sato, K. Ninomiya. Nuclear Isotope Production by Ordinary Muon Capture Reaction. Nuclear Instruments and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment, (March 2020).

Non-Indexed Journal

1. F. Othman, I.H.Hashim, R. Razali, **F. Ibrahim**, F. Soberi. Proton neutron emission model for Muon charge exchange reaction. AIP Conference Proceedings 2319, 080004 (2021); https://doi.org/10.1063/5.0037769, (5 February 2021).

Indexed conference proceedings

1. M I H M Norizan, I H Hashim, **F Ibrahim**, N N A M A Ghani. Statistical Estimation of ideal and realistic muon interactionon Al, Fe, and Cu absorbers. IOP Conference Series: Materials Science and Engineering.(May 2020)
2. **F. Ibrahim**, I.H. Hashim, N. Norjoharuddeen, R. Razali, Liew, C.S. Evaluation of Tracking Reconstruction of Muon-to-Electron Conversion Process Features Through Machine Learning Algorithms. International Nuclear Science, Engineering Conference 2020.

Non-Indexed conference proceedings

1. Marlissa Omar, Aina Hazimah Bahaman, Faridah Aminullah Lubis,Shahrel Ahmad Shuhel Ahmad, **Fahmi Ibrahim**, Siti Norbiha A. Aziz, Fairuz Diyana Ismail,Abd Rahman Bin Tamuri. Perceived Academic Stress Among Students

inUniversiti Teknologi Malaysia. Advances in Social Science, Education and Humanities Research, volume 470. Proceedings of the International Conference on Student and Disable Student Development 2019 (ICoSD 2019), (January 2020)