

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 (NCP), and Data Intensive Computing Centre (DICC) at Universiti Malaya for giving me the opportunity to be Research Assistant at NCP 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 ditentukan 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 klasifikasi isyarat dan latar belakang. Aplikasi pembelajaran mesin dalam zarah fizik sangat berguna dalam meramal sensitiviti eksperimen dan pemprosesan analisa data besar.

TABLE OF CONTENTS

	TITLE	PAGE
	DECLARATION	iii
	DEDICATION	iv
	ACKNOWLEDGEMENT	v
	ABSTRACT	vi
	ABSTRAK	vii
	TABLE OF CONTENTS	viii
	LIST OF TABLES	xii
	LIST OF FIGURES	xiii
	LIST OF ABBREVIATIONS	xx
	LIST OF SYMBOLS	xxii
	LIST OF APPENDICES	xxiv
CHAPTER 1	INTRODUCTION	1
	1.1 Background of Study	1
	1.2 Problem Statement	3
	1.3 Objectives of Study	5
	1.4 Scope of Study	5
	1.5 Significance of Study	6
	1.6 Thesis Outline	7
CHAPTER 2	LITERATURE REVIEW	9
	2.1 Charged Lepton Flavor Violation (CLFV) in Muon Decay	9
	2.2 Experimental Study on CLFV	11
	2.3 COMET Phase-I	12
	2.3.1 Detector Response:Cylindrical Detector (CyDet)	16

2.3.2	Integrated COMET Experimental Data User Software Toolkit (ICEDUST) Framework	23
2.3.3	COMET Geant 4 Simulation	24
2.3.4	COMET Geant4 Geometry	26
2.3.5	Calibration of Simulation Against Ex- perimental Data	27
2.3.6	Data Formats	28
2.3.7	Data Reconstruction by ICEDUST	28
2.3.8	MC Data Production and Distribution	29
2.3.9	MC5 Production File	29
2.4	Machine Learning (ML)	31
2.4.1	Supervised Learning Algorithm: Train- ing and Testing Dataset	33
2.4.2	The Loss Function, Gradient Descent and Regularization	34
2.4.3	Logistic Regression in Classification	36
2.4.4	Gradient Boosted Decision Trees (GBDT)	37
2.5	Machine Learning in Particle Physics for Big Data Analysis	40
2.5.1	Track Finding Classification for COMET Experiment By Using XGBoost Classifier	41
2.5.2	Higgs Boson Decays to Charm-Quark Pairs Using LightGBM by ATLAS Experiment	51
2.5.3	Particle Identification in RICH Detec- tor Using CatBoost by LHCb Experi- ment	52
2.6	Summary of Literature Review	53
CHAPTER 3 METHODOLOGY		55
3.1	Research Flowchart	55

3.2	ICEDUST Simulations	56
3.2.1	SIMG4 Input File (oaRooTracker)	58
3.2.2	Generating Electron Signal At SIMG4 Stage	58
3.2.3	SimHitMerger Stage	59
3.2.4	SimDetectorResponse Stage	60
3.3	Preprocessing Dataset	61
3.3.1	Data Extraction Stage	61
3.3.2	Data Sanitization for Machine Learning Study	63
3.3.3	Selection of Features for Signal and Background	64
3.4	CatBoost Algorithm for Signal Classification	67
3.4.1	Data Testing and Training	68
3.4.2	Performance Evaluation of the CatBoost Classifier	68
CHAPTER 4 RESULT AND DISCUSSION		73
4.1	Observation from MC5A02 Input File	73
4.1.1	Analysis of Muon Stopping Target (MST)	73
4.1.2	Parameters of Bunch Event	74
4.1.3	Detector Hits Timing	76
4.2	CatBoost Classifier as The CDC Local Filter for COMET Phase-I	80
4.2.1	Local and Neighboring Features	82
4.3	Classification Output and Performance of CatBoost Algorithm	87
4.4	Hardware and Computation Timing Metrics	92
4.5	Discussion	93
CHAPTER 5 CONCLUSION		97
5.1	Conclusion	97

5.2	Remarks	98
5.3	Recommendation for Future Works	99

REFERENCES	101
-------------------	------------

LIST OF PUBLICATIONS	121
-----------------------------	------------

LIST OF TABLES

TABLE NO.	TITLE	PAGE
Table 1.1	Summary of MC4o and MC5.	4
Table 2.1	Past experiments on $\mu \rightarrow e$ conversion.	13
Table 2.2	COMET Phase-I experiment for proton beam specifications.	14
Table 2.3	Muon-stopping target configuration.	16
Table 2.4	Main parameters of the CDC.	22
Table 2.5	The summary of MC5.	32
Table 2.6	The summary of previous works and application of GBDT algorithms in particle physics.	54
Table 3.1	The summary of optimization of CTH and CDC for signal and background samples.	64
Table 3.2	The hyperparameters optimized value.	68
Table 4.1	The summarize of machine learning algorithms on the CLFV experiment.	95

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
Figure 2.1	A Feynman diagram of the muon decay.	9
Figure 2.2	An experiment at Brookhaven AGS. The schematic shows a neutrino beam generated through decay in flight of pions. The pions are produced by 15-BeV protons hitting a beryllium target at one end of a 10 feet long straight section. These pions decay to create muon and muon-neutrino pairs. Some of the weakly-interacting neutrinos are able to then pass through over 13.5 meters of steel shielding to hit the 10-ton aluminium spark chamber. The detector watches for the appearance of either muons or electrons resulting from neutrino interactions [28].	10
Figure 2.3	Since the first study of muon decay in 1948, the sensitivity of muonic CLFV searches have improved by 12 orders of magnitude. Note that this plot also shows the evolution of the host nucleus material used in the $\mu^- N \rightarrow e^- N$ searches [37].	11
Figure 2.4	Experimental setup of COMET Phase-I [37].	13
Figure 2.5	The time structure of beam particles [39].	15
Figure 2.6	Schematic diagram of the CyDet system [39].	17
Figure 2.7	The Cherenkov Trigger Hodoscope for Phase-I. The CTH station surround the upstream and downstream ends of the MST as shown in Figure 2.7(a). The stations is located inside the CDC as shown in Figure 2.7(b) [23].	18
Figure 2.8	(right)Organization of Cherenkov and scintillator modules in Cherenkov Trigger Hodoscope for Phase-I. (left)The upstream (Cherenkov) and downstream (scintillator) modules attached with light guide and PMT [23].	19

Figure 2.9	A schematic that highlights the operating principles of a cylindrical drift chamber(CDC). (left)The layout of anode and cathode wires across a simplified wire drift chamber. The anode wires are shown as the filled circles, while the cathodes are the open circles. The solid red line indicates the particle trajectory. (right) A single cell is highlighted in the blue box. The incoming particle ionizes the gas within the cell and creates electrons that accelerate towards the anode wire(refer to green trajectory).	20
Figure 2.10	A demonstration of how stereo angles can recover 3D hit positions. In both figures, the blue layers are rotated around y by a stereo angle θ , while the red layers are rotated around y by θ . This example contains no super-layers. Both images show three colour-coded particles passing through the chamber that leave colour coded hits [23].	21
Figure 2.11	An outline of the ICEDUST framework. The larger blue regions represent parts of the framework that share a common data format, which is specified in the parallelogram [43].	24
Figure 2.12	The package structure of the simulation of ICEDUST.	25
Figure 2.13	Geometry-controller-messenger model class structure [39].	26
Figure 2.14	The concept of sampling world for Phase-I.	30
Figure 2.15	MC5A01 production for upstream about 50 K POT equivalent.	30
Figure 2.16	MC5A02 production for downstream about 50×10^6 POT equivalent.	31
Figure 2.17	An example illustration of Oblivious Decision Tree	40
Figure 2.18	The tree structure of XGBoost.	41
Figure 2.19	The CDC occupancy distribution and an event display from a high occupancy event. Both plots share the color key, denoted in the legend in the occupancy plot to the left. This legend also provides the average number of hits and occupancy for the corresponding entries.	42

Figure 2.20	The signal and background hit charge deposition distribution and corresponding ROC curve. To highlight the separation of the two samples, the distributions are normalized and that the x-axis is displayed on a logarithmic scale. While the detector response yields ADC values in units of e, this plot scales these values by assuming a linear signal amplification in the CDC gas.	44
Figure 2.21	The hit timing and radial distributions for signal and background particles. The timing distribution is taken related to the earliest time of CTH trigger signal in the event. The radial distribution is naturally binned to the number of layers in the CDC.	45
Figure 2.22	The relative timing of the hits to left of the labelled hit. The right plot highlights the dominant zero bin which denotes an empty channel to the left of the labelled hit, while the right shows the distribution in all other bins.	46
Figure 2.23	The normalised classification output distribution of the local GBDT and neighbour GBDT.	47
Figure 2.24	The output of the local and neighbour GBDTs, \hat{y}_{LCL} and \hat{y}_{NGH} , for each hit in the sample event. As in Figure 2.24 (b), the signal hits are colored in pink, while the background hits are blue. The size of the hit is scaled to the output of the corresponding GBDT output, with a full circle denoting a signal-like hit.	48
Figure 2.25	The feature correlation matrix and evaluation metrics. The left plot shows the gain, coverage, and weight of each feature used in the CDC Track Filter. Each metric is normalized to its maximum value across all features. These metrics themselves are defined in the text. The right plot shows the correlations between all features and GBDT scores.	49

Figure 2.26	The final output ROC curves for the baseline ADC classifier, local GBDT, neighbour GBDT, and CDCHF. The left plot measures the performance of the algorithms in terms of their ability to classify signal and background hits. For each implied cut on the left ROC curve, the right ROC curve plots the fraction of signal events that still pass the CDC quality cuts to the average number of background hits that remain in these accepted events [23].	50
Figure 2.27	The tree structure of LightGBM.	52
Figure 2.28	The tree structure of CatBoost.	53
Figure 3.1	Research flowchart.	55
Figure 3.2	An example event from 1.6×10^7 POT per bunch after generating electron signal that coming out of the MST. (a) The hits that came from signal electron tracks in the CDC. (b) The magnitude of momentum of signal distribution which 105 MeV/c at the end of the distribution.	59
Figure 3.3	A bunch train of truth CyDet rates for 20 bunches.	60
Figure 3.4	The work flow conversion of oaEvent file to flat ROOT tree.	62
Figure 3.5	The work flow conversion of oaEvent file to flat ROOT tree.	62
Figure 3.6	An illustration on how the measurement of radial distance from MST.	65
Figure 3.7	The demonstration on how to calculate the angle of hit using local position of hit.	66
Figure 3.8	The illustration of neighbor features are considered, which blue is a wire, yellow is left wire and green is right wire.	66
Figure 3.9	An example of Oblivious Decision Tree using Local features.	67
Figure 3.10	An example of Oblivious Decision Tree using Neighbor features.	67
Figure 3.11	Example of ROC curve (Hoo, 2017).	69

Figure 3.12	Example of Confusion Matrix of binary classification, where True Positives (TP) - prediction and actual both are yes, True Negatives (TN) - prediction is no and actual is yes, False Positives (FP) - prediction is yes and actual is no, and False Negatives (FN) - prediction is no and actual is no.	70
Figure 4.1	(a)The transverse stopping distribution of muons in the Muon Stopping Target disks.This distribution is sum over all disks. (b) The longitudinal stopping distribution of muons in the Muon Stopping Target disks. This plot shows the number of stopped muons per disk, which are perpendicular to the muon beam.	74
Figure 4.2	The effects of the bunch width on the background hit and stopped muon timing in the CDC and CTH. (a) and (b) show the timing distributions for $B_W = 0$ ns, while (c) and (d) plots show these distributions for $B_W = 100$ ns. Note that lines are normalized to one bunch and that the CDC hits do not include the drift time.	75
Figure 4.3	The bunch-train timing structure in the CDC and CTH hits. (a) The full bunch-train window. (b) The plot zooms in on the fiducial time window of the bunch-train event. This plot highlights [500 ns, 1170 ns] as the CTH trigger window. The CDC accepts an additional 180 ns to allow the any signal hits to drift to the wires.	77
Figure 4.4	The overflow of background hits into later fiducial time windows for the CDC and the CTH. The left plots represent the count number of hits from a bunch arriving at $t = 0$ occur within the fiducial time window of the N -th bunch, defined as: [1170 N ns, 500 ns + 1170· ($N + 1$) ns] and denoted as BN . The right plots are the fraction of hits in the first N windows relative to the number of hits in the first 100 windows.	78

Figure 4.5	The distribution of the hit rate from late particles. (a) The hit rates in the CTH separated out by the charge of the particle that leaves the hit. (b) The hit rate separated out by the charge of the parent of the particle that made the hit.	79
Figure 4.6	The distribution of the hit rate from late particles. (a) The hit rates in the CDC separated out by the charge of the particle that leaves the hit. (b) The hit rate separated out by the charge of the parent of the particle that made the hit.	80
Figure 4.7	The CDC occupancy and the CTH trigger acceptance over multiple bunches. The histogram counts the hits in each detector using the left axes. The occupancy and trigger acceptance are plotted with the blue line and correspond to the right axis.	81
Figure 4.8	The signal and background hit charge deposition distribution.	82
Figure 4.9	The radial distribution for signal and background particles. The radial distribution is naturally binned to the number of layers in the CDC.	83
Figure 4.10	The hit timing distribution for signal and background particles. The timing distribution is taken relative to the time of the earliest CTH trigger signal in the event.	83
Figure 4.11	The angle of hit, ϕ from x -axis distribution for signal and background particles.	84
Figure 4.12	The event display of the CDC coloured by and background hits in a mixed sample truth label.	85
Figure 4.13	The neighboring charge deposit distributions are (a) the signal and background particle deposit distribution in the left wire, q_L , and (b) the signal and background particle deposit distribution in the right wire, q_R .	86
Figure 4.14	The following are the neighboring distributions of signal and background particle timings: (a) the relative hit timing distribution for signal and background particles on the left wires, t_{RL} , and (b) the relative hit timing distribution for signal and background particles on the right wires, t_{RR} .	86

Figure 4.15	The normalised classification output distribution of the (a)Local CatBoost and (b)Neighbor CatBoost.	87
Figure 4.16	The confusion matrix of (a)Local CatBoost and (b)Neighbor CatBoost for default threshold.	88
Figure 4.17	The final output ROC curves for the baseline time relative of hit to trigger signal classifier, time relative of hit to trigger signal classifier and radial distance of MST, Local CatBoost, Neighbour CatBoost. This result of ROC is for default threshold ($y=0.5$) and the points on the lines represents the trade-off point between background rejection efficiency and signal retention efficiency.	89
Figure 4.18	The final output ROC curves for the baseline time relative of hit to trigger signal classifier, time relative of hit to trigger signal classifier and radial distance of MST, Local CatBoost, Neighbour CatBoost. This result of ROC is for optimized threshold and the points on the lines represents the trade-off point between background rejection efficiency and signal retention efficiency.	90
Figure 4.19	The output of the (a)Local and (b)Neighbour CatBoosts for each hit in the sample event. The signal hits are colored in red, while the background hits are blue.	91
Figure 4.20	The feature correlation matrix. The plot shows the correlations between all features and Local CatBoost, y_{local} and Neighbor CatBoost, $y_{neighbor}$ scores.	92
Figure 4.21	The feature importance of each feature based on PredictionValuesChange metric.	93
Figure A.1	Total energy deposit in CTH according PID.	113
Figure A.2	Total energy deposit in CDC according PID.	113

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 Analytics
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 Scherrer Institute
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

APPENDIX	TITLE	PAGE
Appendix A	Energy Deposit in CTH and CDC	113
Appendix B	LogLoss Against Number of Trees	114
Appendix C	Accuracy Against Number of Trees	115
Appendix D	Local CatBoost & Neighbor CatBoost According Trigger Track ID	116

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:

$$E_{\mu e} = m_\mu - B_\mu - E_{recoil} \quad (1.1)$$

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 $B(\mu^- + Au \rightarrow e^- + Au) \leq 7 \times 10^{-13}$) 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 AICap 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 AICap 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 AICap 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 AICap. 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 Phase-I	Yes	No
Sampling World Physic list	No	Yes
	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 AICap Experiment. *arXiv preprint arXiv:1809.10122*, 2018.
19. Wyttenbach, A., Baertschi, P., Bajo, S., Hadermann, J., Junker, K., Katcoff, S., Hermes, E. and Pruyss, 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://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 Almpandis, 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 $\mu^+ \rightarrow e^+ \gamma$ 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 $\mu^+ \rightarrow e^+ \gamma$. *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 $\mu \rightarrow eee$. *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 $\mu^+ \rightarrow e^+ e^+ e^-$. *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 $\mu \rightarrow e^+$ 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 \rightarrow 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 $\mu \rightarrow e$ 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 $\mu \rightarrow e$ 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., Dvornický, 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 Rönvaldsson, 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 Unlimited 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. Saarimäki, 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)