Using Neural Networks to Analyse Hadronic Vacuum Polarisation Contribution to Muon g-2

A REPORT

submitted in partial fulfillment of the requirements

for the award of

5 Year BS-MS Dual Degree

in

Physics

by

Sidharth Gupta

(19MS067)

Under the guidance of

Dr. Pere Masjuan

Department of Physical Sciences
Indian Institute of Science Education and Research, Kolkata

April 2024

DECLARATION

Date: 22/04/2024

I, Mr. Sidharth Gupta Registration No. 19MS067 dated 31/07/2019, a student of the Department of Physical Science of the 5 Year BS-MS Dual Degree Programme of IISER Kolkata, hereby declare that this thesis is my own work and, to the best of my knowledge, it neither contains materials previously published or written by any other person, nor has it been submitted for any degree/diploma or any other academic award anywhere before. I have used the originality checking service to prevent inappropriate copying.

I also declare that all copyrighted material incorporated into this thesis is in compliance with the Indian Copyright Act, 1957 (amended in 2012) and that I have received written permission from the copyright owners for my use of their work.

I hereby grant permission to IISER Kolkata to store the thesis in a database which can be accessed by others.

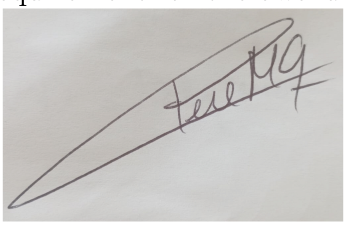
Sidharth Gupta

Sidharth Gupta
Department of Physical Sciences
Indian Institute of Science Education and Research Kolkata
Mohanpur 741246, West Bengal, India

CERTIFICATE

Date: 16/05/2024

This is to certify that the thesis titled "Using Neural Networks to analyse Hadronic Vacuum Polarisation Contribution to muon g-2" submitted by Mr. Sidharth Gupta Registration No. 19MS067 dated 16/05/2024, a student of the Department of Physical Sciences of the 5 Year BS-MS Dual Degree Programme of IISER Kolkata, is based upon his/her own research work under my supervision. I also certify,to the best of my knowledge, that neither the thesis nor any part of it has been submitted for any degree/diploma or any other academic award anywhere before. In my opinion, the thesis fulfils the requirement for the award of the 5 year BS-MS Dual Degree.



Dr. Pere Masjuan
Associate Professor & Coordinator of the Degree in Physics & Head of the Theory Group
Departament de Física & Institut de Física d'Altes Energies
Grup de Física Teòrica
Campus de la UAB, 08193, Bellaterra
Cerdanyola del Vallès, Barcelona, Spain

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have contributed to the completion of this thesis.

First and foremost, I am deeply thankful to my thesis supervisor, Dr. Pere Masjuan, for his invaluable guidance, support, and encouragement throughout the research process. His expertise, insightful feedback, and unwavering dedication have been instrumental in shaping this work.

I extend my heartfelt appreciation to IISER Kolkata for providing the necessary resources and institutional support.

I am grateful to Professor Supratim Sengupta for coordinating with my supervisor for the evaluation of this work.

I would like to acknowledge the contributions of my colleagues, friends, and family members who have offered their encouragement, understanding, and moral support throughout this journey. Their belief in my abilities and unwavering support have been a source of strength and motivation.

Thank you to all who have played a part, no matter how big or small, in bringing this thesis to fruition.

Sidharth Gupta

CONTENTS

1. INTRODUCTION

The recent Muon g-2 anomaly has reignited the debate in particle physics, challenging the Standard Model's ability to fully explain experimental observations.[3] Supporters of the Standard Model argue that it has successfully predicted and explained numerous phenomena, and that the anomaly may be resolved through refinements within its framework. Conversely, proponents of Beyond Standard Model theories seize upon the anomaly as evidence of the model's incompleteness, advocating for new physics beyond the established paradigm[4]. While the anomaly has yet to be conclusively explained, it underscores the importance of continued experimentation and theoretical exploration to unravel the mysteries of particle physics.

A muon possesses an intrinsic magnet akin to a miniature bar magnet and angular momentum known as spin, analogous to a spinning top. The muon's gyromagnetic ratio "g" is determined by the strength of this magnet and the speed of its rotation. The Muon g-2 experiment is named after the minute deviation in the muon's gyromagnetic ratio "g", which differs slightly, approximately by 0.1 percent, from the anticipated value of 2[5]. This discrepancy, often termed the anomalous magnetic moment of the muon, $a_\mu = (g-2)_\mu/2$ prompts further investigation into the underlying physics beyond the anticipated predictions. The results of this experiment hold a lot of potential to explain this or make other conclusions.

For muon g-2 experiment, hadronic vacuum polarization refers to the influence of virtual hadronic particles, primarily pions, on the muon's anomalous magnetic moment. As the muon interacts with these virtual particles in the vacuum, its magnetic properties are altered. Precisely understanding this effect is crucial for interpreting experimental data, as it constitutes a significant source of uncertainty in the theoretical prediction of the muon's magnetic moment. Reducing this uncertainty is essential for uncovering potential deviations from the Standard Model and exploring new physics.

This thesis will solely focus on the hadronic vacuum polarization contribution towards the anomalous magnetic moment of the muon, $a_\mu^{had,VP}$. The reason for focusing on this specific contribution is that the discrepancy that $a_\mu^{had,VP}$ holds is 3.5σ [1] or higher. One aspect of looking at this discrepancy is that different collaborations have measured this value to be different (some very far from other) and so to consider them all brings a huge discrepancy in the final result. The aim of this work is to use the data from these organisations and make a model that learns from all of them, unbiased, and gives us a combined dataset/value. There are more conclusions that can be made from the model that shall be discussed in subsequent chapters.

We have experimented and made a model using Neural Networks and some of its many libraries provided by TensorFlow and Keras and other Python libraries. More about the structure of model, in detail, will be discussed in subsequent chapters. After putting forward the problem and how the model will tackle the problem the results obtained from it and the conclusions derived from them will be discussed.

2. WHY HADRONIC VACUUM POLARISATION?

The Muon g-2 experiment delves into the intricacies of the muon, a fundamental particle with intriguing magnetic properties and angular momentum akin to a spinning top. At its core lies the muon's gyromagnetic ratio "g", a parameter revealing deviations from the anticipated value of 2, termed the anomalous magnetic moment[4]. This discrepancy, around 0.1 percent, prompts a meticulous investigation into the muon's behavior within magnetic fields.

When subjected to a magnetic field, the muon's internal magnetism interacts, causing a phenomenon called precession. This precession rate, influenced by the muon's g-value, undergoes modification due to transient particle interactions within the vacuum, thereby altering the g-2 term. The precision demanded by modern physics calls for measurements of g-2 accurate to an astonishing 140 parts per billion, equivalent to scrutinizing the length of a football field down to a tenth of a human hair's thickness.[5]

The renowned Standard Model of particle physics furnishes a precise forecast for the muon g-2, accurate to 400 parts per billion. In contrast, the Fermilab Muon g-2 experiment aspires to surpass this prediction with quadruple the precision of its predecessor at Brookhaven National Laboratory[6]. This heightened accuracy aims to scrutinize the subtle discrepancy between experimental observations and the Standard Model's predictions, potentially heralding groundbreaking revelations in particle physics.

Central to this endeavor is the exploration of the hadronic vacuum polarization contribution to g-2. While perturbative Quantum Chromodynamics (pQCD) offers insights into high-energy phenomena, its applicability wanes in the low-energy realm where quarks are confined within hadrons. The intricacies of QCD at low energies, characterized by confinement and chiral symmetry breaking, elude precise prediction through conventional pQCD methods. It's noteworthy that the running strong coupling constant, $\alpha_s(s)$, can only be reliably trusted beyond approximately 2 GeV, away from thresholds and resonances[2].

As a result, a semi-phenomenological approach becomes imperative, leveraging dispersion relations, the optical theorem, and empirical data to unravel the enigmatic behavior of hadronic contributions to g-2 [2]. This multifaceted investigation not only promises to refine our understanding of fundamental particles but also holds the potential to unveil new physics beyond the established paradigms of the Standard Model, paving the way for transformative advancements in the field.

From the various channels contributing towards $a_\mu^{had,VP}$, this work shall be focused on the datasets for $e^+e^- \longrightarrow \pi^+\pi^-$ channel. The choice of working on these datasets (provided by different collaborations) is because this channel adds the majority of the contribution(upto 75%) and also the uncertainty in the final calculation. One way of looking at why this specific channel has a huge uncertainty is that different collaborations providing different results. Is this a problem of disagreement, different ways of calculation or something else we are overlooking?

The datasets that have been used for training the Neural Network model are provided by KLOE, BaBar and CMD-3 [7][8][9]. These collaborations have the bare cross-section measured for the channel we're focusing on.

The problem is not yet completely put since the mathematics that is involved in the calculation of $a_\mu^{had,VP}$ had been left out. What matters here is that we do not require specific data points or a data range, but, as will be shown next, we require the integral under the curve with some kernel and normalization factor that shall be put forward next.

The $O(\alpha^2)$ contributions to $a_\mu^{had,VP}$ can be calculated through equation (2.1) given below[2]:

$$a_\mu^{(4)}(vap, had) = \Big(\frac{\alpha m_\mu}{3\pi}\Big)^2 \left(\int_{m_{\pi^0}^2}^{E_{cut}^2} ds \frac{R_{had}^{data}(s)\hat{K}(s)}{s^2} + \int_{E_{cut}^2}^{\infty} ds \frac{R_{had}^{pQCD}(s)\hat{K}(s)}{s^2} \right) \tag{2.1}$$

(vap, had) stands for hadronic vacuum polarisation.
The Kernel $K(s)$, written in terms of the following,

$$x = \frac{1-\beta_\mu}{1+\beta_\mu}, \qquad \beta_\mu = \sqrt{1 - 4m_\mu^2/s} \tag{2.2}$$

is given by:

$$K(s) = \frac{x^2}{2}(2-x^2) + \frac{(1+x^2)(1+x)^2}{x^2}\left(\ln(1+x) - x + \frac{x^2}{2}\right) + \frac{(1+x)}{(1-x)}x^2\ln(x) \tag{2.3}$$

and $\hat{K}(s) = \frac{3s}{m_\mu^2}K(s)$, varies slowly in the interested range of integration, increasing monotonically from 0.63 at $s = 4m_\pi^2$ to 1 at $s = \infty$,
and $R_{had}^{data}(s) = \sigma / \frac{4\pi\alpha(s)^2}{s}$

Our task is limited to the calculation of the integral on the left in equation (2.1), simply using the provided result. Different collaborations may have used different ways of calculation also.

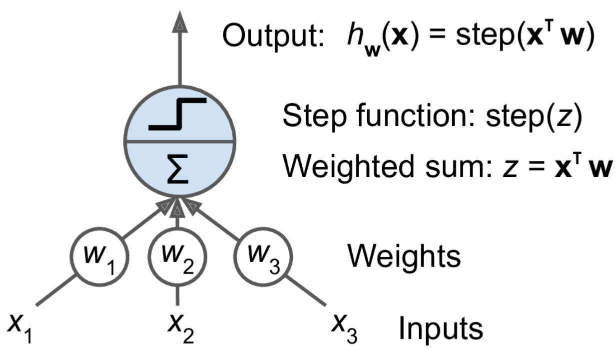
This work tackles this problem by making a model that takes datasets from these collaborations and learns from them unbiased, providing another “combined dataset” that gives results agreeing with each dataset. The approach for making this model is using Neural Networks to build a complex model that can learn every small trend and details, more of which shall be discussed in next chapter.

To summarise the problem: We need to make a model that accepts original datasets from different collaborations and provides us a combined dataset that we shall further use to calculate the integral on the left in equation (2.1). We're doing this for the $e^+e^- \longrightarrow \pi^+\pi^-$ channel because this channel is the one that is responsible for most of the uncertainty in the final calculations.

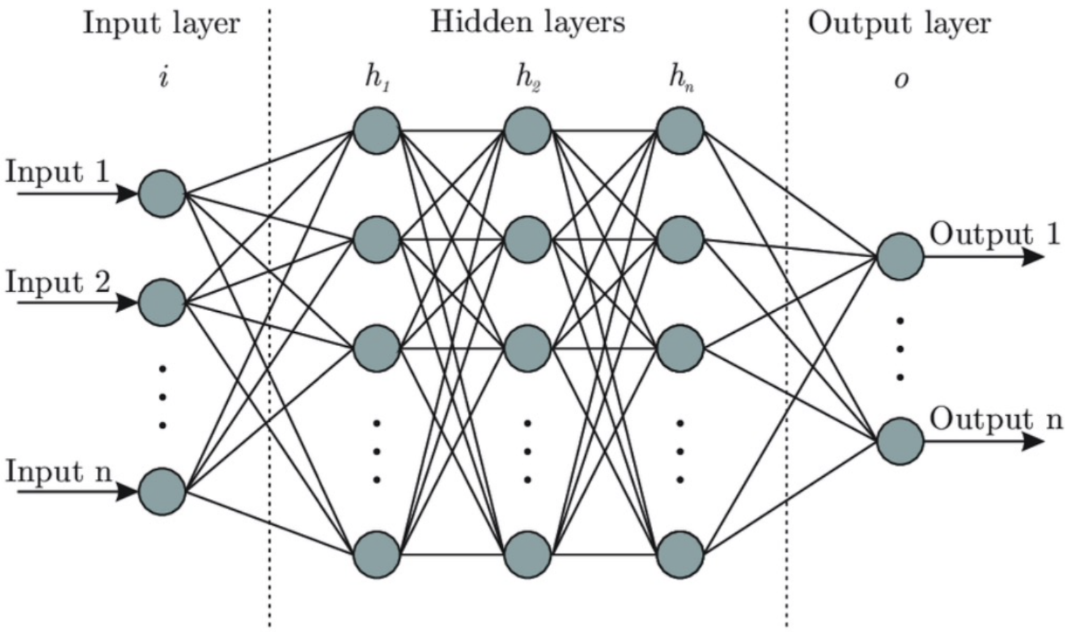
3. NEURAL NETWORK MODEL

Machine learning is a field of artificial intelligence that empowers computers to learn from data and improve their performance over time without being explicitly programmed. It involves the development of algorithms and models that enable machines to recognize patterns, make predictions, and optimize decision-making based on input data. The concept of Artificial Neural Networks (ANNs) draws inspiration from nature, where various inventions have been inspired by the inherent design of living organisms. In particular, ANNs take inspiration from the intricate networks of biological neurons in the human brain.

A Perceptron is one of the simple ANN architectures as shown in the image below, it is based on an artificial neuron known as Threshold Logic Unit (TLUs) also called Linear Threshold Unit (LTUs). Instead of being binary values, the inputs and outputs are numbers and every input connection has an associated weight with it. The TLU has a simple job computing the weighted sum of the inputs and then apply a step function, more commonly known as activation function, to that sum and hence giving the output[10].



So far, this looks very simple and easy to even be done by hand but our aim is to build a model that learns complex patterns and details in the data from which we aim to draw conclusions.
A general pictorial representation of a Deep Neural Network is as follows[10]:



Here we control the number of "neurons" in each layer and also the number of hidden layers, each neuron being connected to each neuron in its previous layer, making the architecture very complex and having a lot of weighted connections.

The data sets we have have energy values as x axis and bare cross-section as the y axis. The model takes x values as input and gives output using random initialized weights, these weights are updated using the difference between the actual y value and the predicted value of the model. There are various types of loss functions, optimization techniques and other hyperparameters that can be changed to make the model learn better and give better output. The code to the model that I made is attached in the appendix for reference. Having a lot of options of hyperparameters and other optimization techniques provides a huge number of possibilities in which a model can be made to learn. The theory above is just to explain the basic working of neural networks, in reality there is a lot more to neural networks in the realm of Deep Learning.

The data sets form each organisation overlap in a specific energy region and hence we do not have a lot of data to train the model. A good number of data points to train lie in the range from thousands to millions or even billions, but we only have around 385 data points to work with making it really challenging to make a model that learns fast and also does not over-fit the data. Nevertheless, the model made by the end of this work is giving intriguing results which shall be put forward and discuss in the next chapter. Making this model was not an overnight job as explained in previous paragraph, there a lot of different ways a model can be changed and made better or worse. It involves doing a lot of hit and trial runs, also with gradually making the model better and better, it took longer for each run and the run that gave the final results took more than three hours, sometimes even six.

The model architecture, denoted as 'build_model' in Appendix 1, is designed to process input data through parallel pathways before merging them and producing the final output. The model comprises the following components:

1. Input Layers: Two input layers ('input1') are defined with a specified shape ('input_shape1'). These serve as the entry points for the data into the neural network.

2. Parallel Pathways (Branches): The model features two parallel pathways, denoted as 'l1' and 'l2', which process the input data separately. Each pathway consists of:
- Two densely connected layers with rectified linear unit (ReLU) activation functions. These layers serve to extract and transform features from the in-

put data.
- Batch normalization layers applied after each dense layer, aiding in stabilizing and accelerating the training process.
- Dropout layers, set to a dropout rate of 0.2, which randomly deactivate a fraction of neurons during training to prevent overfitting.

3. Concatenation: The outputs of the two pathways ('l1' and 'l2') are concatenated together, combining the feature representations learned by each pathway into a single representation.

4. Merged Layers: Following concatenation, the merged features undergo further processing through two additional densely connected layers, similar to those in the parallel pathways. These layers continue to refine the feature representations.

5. Final Output Layer: The processed features are passed through a final densely connected output layer, which produces the model's output. The number of neurons in this layer corresponds to the specified 'output_shape'.

6. Model Compilation: The model is compiled using suitable loss functions (huber here) and optimization algorithms (Adam here) to prepare it for training and evaluation.

4. RESULTS

The data sets available are from three collaborations; KLOE, BaBar and CMD-3 [7][8][9]. The plots for each dataset are as shown in Fig. 4.1. As mentioned in Chapter 2, our task is to provide a "combined dataset" from which we may calculate the integral on the left in equation (2.1).

Even though the plots in Fig. 4.1 look all very close to each other, when calculations are made to get the aforementioned integral and the value of hadronic vacuum polarisation contribution within a_μ, the uncertainty is high for this channel ($e^+e^- \longrightarrow \pi^+\pi^-$), a lot more than that for other channels, which is the reason why this work is focused on the datasets from this channel.

As promising as the plot for the dataset from the model in looks Fig. 4.2, it is not yet what we had aimed to achieve, we can see in the plot to the right the difference between the predicted values from my model and actual values from each dataset. A good fit would have given a total of zero error in each energy region/range and also the errors would have been a little smaller in value. For example in the energy region from 0.73 to 0.77 the model also gives a, somewhat, better result than that in other regions since the sum of errors from each dataset comes out to be close to zero (again, not very close).

The main problem that the model is facing is that the total data that has been used to fit the model is very scarce, in fact, there are only 385 data points in the region where all three data sets overlap, not claiming that this model is a perfect model. There are always other approaches to make a model better that may be able to learn better and faster with limited number of data points.

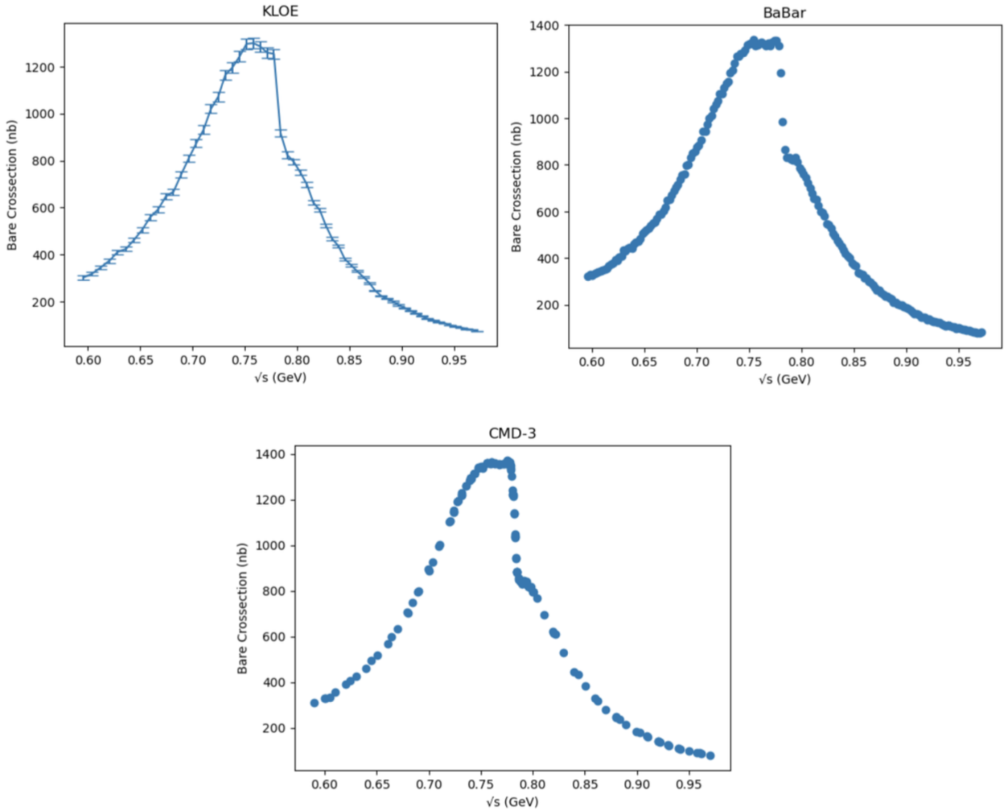


Fig. 4.1: These are the plots for the datasets from respective collaborations

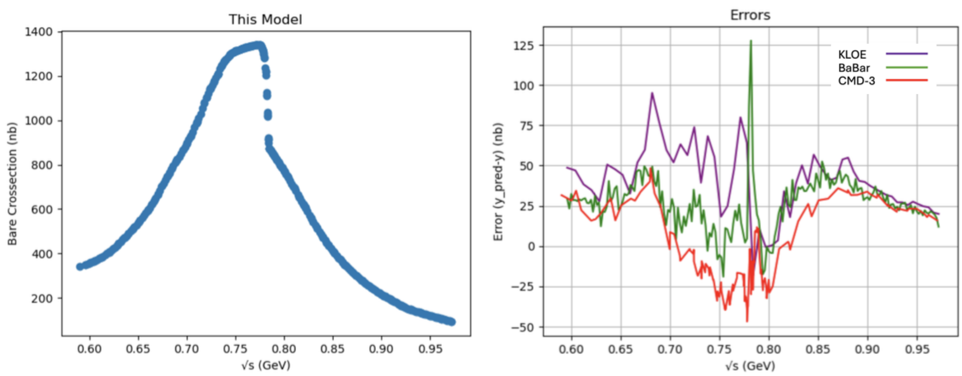


Fig. 4.2: On the left is the plot for the dataset that the model for this work gives and on the right is the plot for difference between the predicted values by the model and the actual data from the three collaborations.

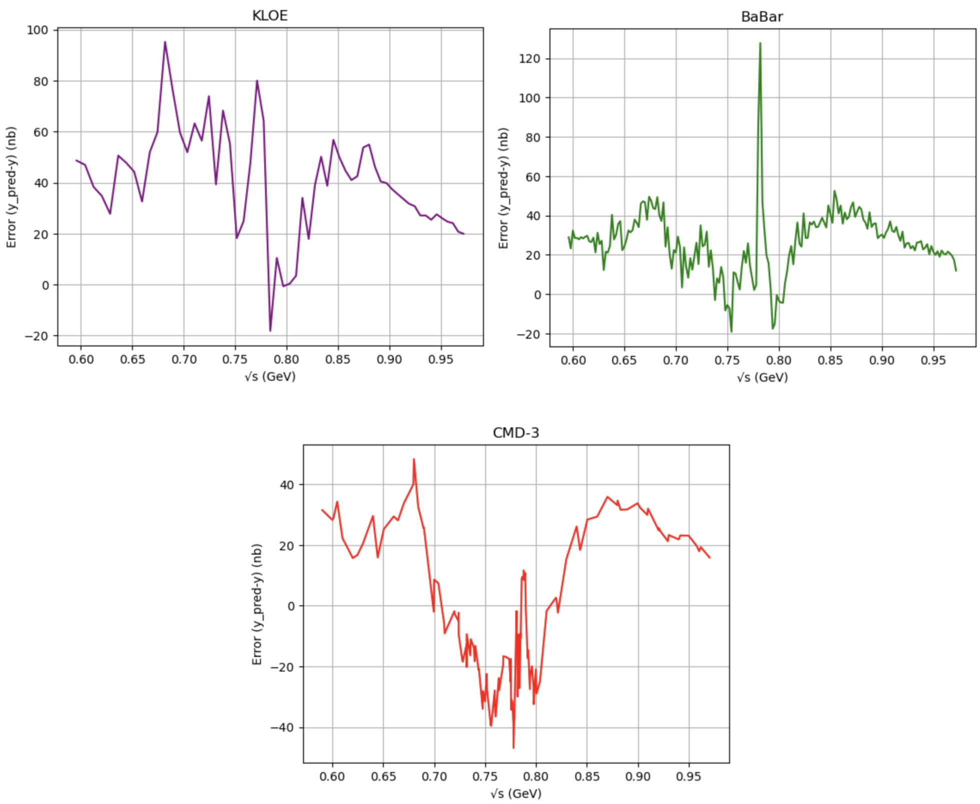


Fig. 4.3: These are the individual error plots (Predicted values - Actual values) for the three collaborations

We can see, clearly, in energy range 0.72 to 0.82, the error for KLOE lies in far positive, that for BaBar lies very close to zero and the same for CMD-3 lies, almost all, in negative. Analytically, this is the kind of result that was the aim for this work to achieve, only better than what has been achieved here. Nevertheless, this work shall be seen as a proof of concept that if we have more data or a better model, that can provide us with a better result, the uncertainty in the final result can then be reduced.

For $\sqrt{s}$ in the range 0.6 to 0.95 (almost the complete range), the integral under each error curve using trapezoidal rule comes to be 14.39 for KLOE, 9.46 for BaBar, 4.16 for CMD-3. This is not the desired result as the sum of these three values should lie close to zero which is not the case when checked for the complete range (0.6,0.95). But, when we check in the $\sqrt{s}$ range from 0.76 to 0.82, the integral was calculated to be 1.16 for KLOE, 1.1 for BaBar and -0.93 for CMD-3. This is, again, not the perfect result but way better than the rest of the regions.

The trapezoidal integral under the curve for datasets from each collaboration and the dataset from this work are;

KLOE	211.85
BaBar	217.52
CMD-3	224.62
This Model	229.5

Analytically speaking, the integral from this model should lie in between the values of the other three.

Now, calculating the integral on the left side in equation (2.1) using the dataset from the model and also from the three collaborations, we get;

	$a_\mu^{(4)}(vap, had)$
KLOE	382.32×10^{-10}
BaBar	392.75×10^{-10}
CMD-3	406.06×10^{-10}
This Model	414.64×10^{-10}

All these calculations were done using the integral on the left in equation (2.1), taking $\alpha(s) = 1/137$ in the interested range of the data (0.6,0.95), we can see clearly from this result that the model in this work is not perfect, but, on the contrary, it clearly provides the motivation to use Neural Networks to build better models which can fit these datasets together better than what has been done in this work. It also provides motivation to get more experimental data to help the learning of Neural Network models learn from a huge dataset and not just 385 data points.

5. CONCLUSION

In conclusion, the Muon g-2 anomaly continues to challenge the Standard Model's ability to fully explain experimental observations. This thesis presented a novel approach to merging data from multiple sources using a neural network model. While the initial results are promising, particularly in the 0.73 to 0.77 energy region where the sum of errors from each dataset is close to zero, there are still regions where the model's predictions deviate from the actual values.

This highlights the limitations of training a complex model with a relatively small dataset (385 data points). Future work could involve expanding the data used for training, potentially by incorporating data from future experiments or utilizing techniques to generate synthetic data. Additionally, exploring different neural network architectures or training methodologies might yield further improvements in the model's accuracy.

Despite these limitations, this thesis demonstrates the potential of applying neural networks to combine data from disparate sources and contribute to our understanding of complex phenomena like the Muon g-2 anomaly. Further development and refinement of this approach, coupled with advancements in data acquisition and processing, hold significant promise for future research in particle physics.

Appendices

I Neural Network Model

Written below is the code of my model. To fit this model the data needs to be reshaped and the input shape and output shape needs to be specified. After using different optimizers, Adam worked the best with learning rate = 0.0000009.

I built/changed the model many times to obtain the one below, this worked the best for me.

When using the fit method, sending a batch of 385 per epoch and 1800000 epochs worked the best after many different trial runs. After model finished fitting (taking 5-6 hours), .predict() method was used to obtain the data in the desired x-range which was further compared to actual data sets as shown in Results chapter.

```
def build_model(input_shape1, output_shape):
    input1 = keras.layers.Input(shape=input_shape1, name='input1')
    l1 = keras.layers.Dense(128, activation='relu')(input1)
    l1 = keras.layers.BatchNormalization()(l1)
    l1 = keras.layers.Dropout(0.2)(l1)
    l1 = keras.layers.Dense(64, activation='relu')(l1)
    l1 = keras.layers.BatchNormalization()(l1)
    l1 = keras.layers.Dropout(0.2)(l1)
    l2 = keras.layers.Dense(128, activation='relu')(input1)
    l2 = keras.layers.BatchNormalization()(l2)
    l2 = keras.layers.Dropout(0.2)(l2)
    l2 = keras.layers.Dense(64, activation='relu')(l2)
    l2 = keras.layers.BatchNormalization()(l2)
    l2 = keras.layers.Dropout(0.2)(l2)
    merged = keras.layers.concatenate([l1, l2])
    merged = keras.layers.Dense(128, activation='relu')(merged)
    merged = keras.layers.BatchNormalization()(merged)
```

```
merged = keras.layers.Dropout(0.2)(merged)
merged = keras.layers.Dense(64, activation='relu')(merged)
merged = keras.layers.BatchNormalization()(merged)
merged = keras.layers.Dropout(0.2)(merged)
output = keras.layers.Dense(output_shape, name='output1')(merged)
model = keras.models.Model(inputs=[input1], outputs=output)
return model
```
[10][11][12]

This model is only the one that worked best for me. Better models can always be made and other methods could also enhance the learning of models.

BIBLIOGRAPHY

1. Keshavarzi, Alexander, et al. ‘Muon $g-2$ and α (M Z 2): A New Data-Based Analysis’. Physical Review D, vol. 97, no. 11, June 2018, p. 114025. DOI.org (Crossref), https://doi.org/10.1103/PhysRevD.97.114025.

2. Jegerlehner, Fred, and Andreas Nyffeler. ‘The Muon $G-2$’. Physics Reports, vol. 477, no. 1–3, June 2009, pp. 1–110. DOI.org (Crossref), https://doi.org/10.1016/j.physrep.2009.04.003.

3. https://muon-g-2.fnal.gov/

4. https://muon-g-2.fnal.gov/the-physics-of-g-2.html

5. https://muon-g-2.fnal.gov/how-does-muon-g-2-work.html

6. https://muon-g-2.fnal.gov/key-contribution-from-brookhaven.html

7. @articleCMD-3:2023alj, author = “Ignatov, F. V. and others”, collaboration = “CMD-3”, title = “Measurement of the $e^+e^- \to \pi^+\pi^-$ cross section from threshold to 1.2 GeV with the CMD-3 detector”, eprint = “2302.08834”, archivePrefix = “arXiv”, primaryClass = “hep-ex”, month = “2”, year = “2023”

8. @articleKLOE:2012anl, author = “Babusci, D. and others”, collaboration = “KLOE”, title = “Precision measurement of $\sigma(e^+e^- \to \pi^+\pi^-\gamma)/\sigma(e^+e^- \to \mu^+\mu^-\gamma)$ and determination of the $\pi^+\pi^-$ contribution to the muon anomaly with the KLOE detector”, eprint = “1212.4524”, archivePrefix = “arXiv”, primaryClass = “hep-ex”, doi = “10.1016/j.physletb.2013.02.029”, journal = “Phys. Lett. B”, volume = “720”, pages = “336–343”, year = “2013”

9. @articleBaBar:2012bdw, author = “Lees, J. P. and others”, collaboration = “BaBar”, title = “Precise Measurement of the $e^+e^- \rightarrow \pi^+\pi^-(\gamma)$ Cross Section with the Initial-State Radiation Method at BABAR”, eprint = “1205.2228”, archivePrefix = “arXiv”, primaryClass = “hep-ex”, reportNumber = “BABAR-PUB-12-003”, doi = “10.1103/PhysRevD.86.032013”, journal = “Phys. Rev. D”, volume = “86”, pages = “032013”, year = “2012”

10. Aurelien Geron - Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems-O’Reilly Media, Inc. (2023).pdf

11. https://scikit-learn.org/0.21/documentation.html

12. https://www.tensorflow.org/api_docs/python/tf/all_symbols