**Technical University of Applied Sciences Würzburg-Schweinfurt (THWS)**

Faculty of Computer Science and Business Information Systems

## Master Thesis

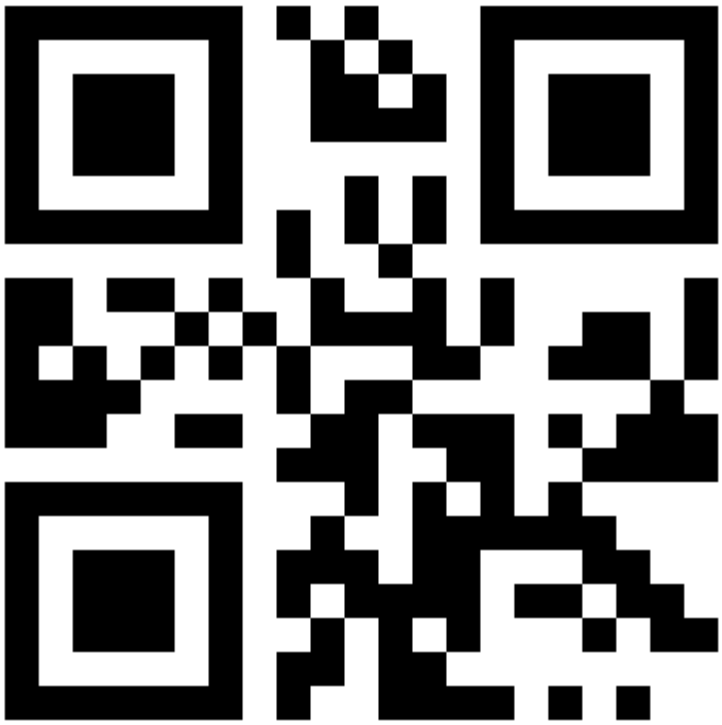
# Electrical Engine Efficiency Prediction Bypassing PDE Simulator

**Submitted to the Technical University of Applied Sciences Würzburg-Schweinfurt in the Faculty of Computer Science and Business Information Systems to complete a course of studies in Master of Artificial Intelligence**

Lilly Abraham
K64889

Submitted on: 11.12.2024

Initial examiner: Prof. Dr. Magda Gregorova
Secondary examiner: Prof. Gracia Herranz Mercedes



# Abstract

The thesis explores an approach to predict Key Performance Indicators(KPI)s of topology invariant Interior Permanent Magnet Synchronous Magnets (IPSM) Electric Motors from its geometric, physical and simulation parameters.
The KPIs to be predicted are plots on Efficiency grid(3D) and Torque curve(2D).
We aim to first parameterize the Electric Motor design such that it is feasible to convert into a tabular representation.
Next, we would create a table with relevant attributes and design a Multi Linear Perceptron(MLP) with the table as input and the plots in its numerical format as vectors of values.
Additionally we conduct a study to model the task as its graph representation and use Graph Neural Networks(GNN) to predict the KPIs.
Subsequently we will regularize the loss functions in a way that would smoothen out the plot curves of the prediction values.
Then we will evaluate the predictions with the test target values by experimenting with various hyperparameter tuning settings and as a baseline with the average of the parameters.
Finally we will enable the KPI's plot visualisation in a manner presentable to the client Valeo(Automaker Company).

# Acknowledgement

I would like to thank my supervisor Prof. Dr. Magda Gregorova for her guidance and support throughout the course of this thesis. Her dedication and commitment to our work has been inspiring to me especially on how we transformed statistical math into modelling that I might have developed a new love for academia.
I would also like to express my sincere gratitude to Valeo for providing us with the dataset. Special thanks are in order to Daniel and Leo for sharing valuable insights of the data from an electromechanical standpoint and for giving me a detailed understanding of my task.
I owe my Mother and my Siblings my accomplishments. They have been very instrumental in making it possible for me to pursue a Master's degree outside my home country and for the endless support thoughout.
Finally, I humble myself before God Almighty for all his blessings and for giving me the strength to persevere and bring my dreams to fruition.

# Contents

# Abbreviations

| | |
|---|---|
| **GNN** | Graph Neural Network |
| **MLP** | Multi Linear Perceptron |
| **KPI** | Key Performance Indicator |
| **EM** | Electric Motor |
| **FEM** | Finite Element Method |
| **CNN** | Convolution Neural Network |
| **2D** | 2 Dimension |
| **3D** | 3 Dimension |
| **MSE** | Mean Squared Error |
| **RMSE** | Root Mean Squared Error |
| **NaN** | Not a Number |
| **ReLU** | Rectified Linear Unit |
| **GPU** | Graphics Processing Unit |
| **MP** | Message Passing |

# Chapter 1

# Introduction

In the design of electric motors, vast amounts of data are generated to determine which design of an EM fits best to KPIs.
KPIs of an Electric Motor are essential to judge the performance of the motor before it is manufactured.
Traditionally these KPIs are inferred from a description of an EM design via a Finite Element Method (FEM) approximating the solutions of the Maxwell's equations.
This process, though well established in the EM design, is very time consuming and does not allow for high-throughput engine design optimization.
The actual engine data of Valeo is used here as the dataset comprising of multiple variant designs of the Double-V topology.

The 3 motor topologies manufactured by Valeo are as below:

1. Single V Magnet - Consists of a single V magnet shown in Figure 1.1.
2. Double V Magnet - Consists double V magnets shown in Figure 1.2.
3. Nabla Magnet - Consists of a single V Magnet and a delta magnet shown in Figure 1.3.

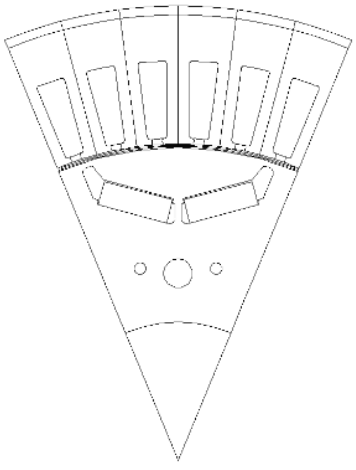


Figure 1.1: V1 Magnet
(Source: Valeo)

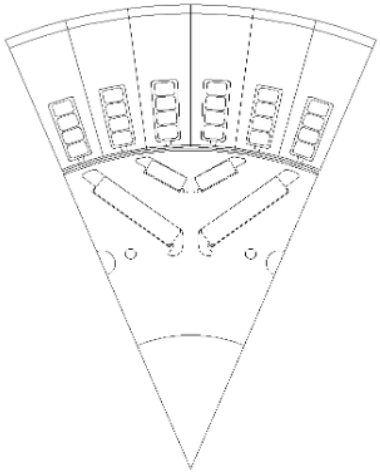


Figure 1.2: V2 Magnet
(Source:Valeo)

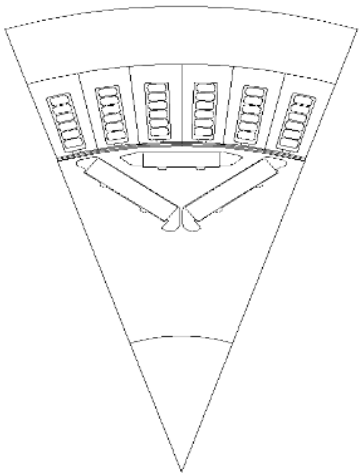


Figure 1.3: Nabla Magnet
(Source:Valeo)

Figure 1.4 and Figure 1.5 gives a glimpse of the EM KPIs to be predicted.

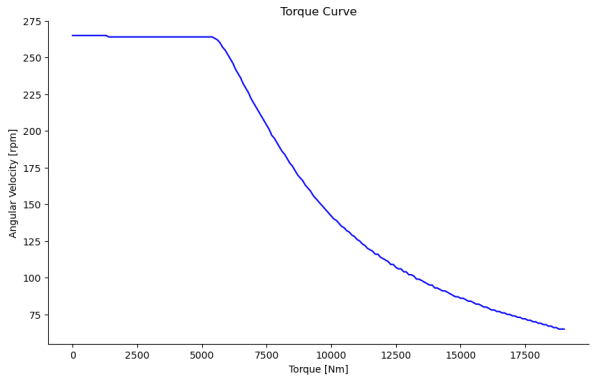



Figure 1.4: Torque Curve

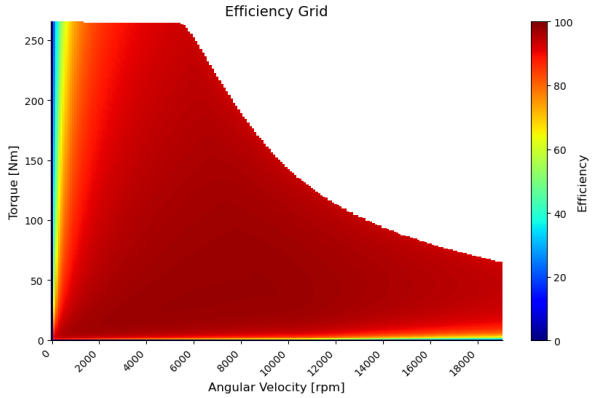



Figure 1.5: Efficiency Grid

The current pipeline to predict the KPIs of different EM design variants is to create a design mesh from the parametric description of the motor with Matlab. Multiple FEM simulations which are by nature Partial Differential Equations(PDE) is done on this mesh which is then post processed and the intermediary outputs are forwarded to the Motor Builder. Several Motor builder settings are then adjusted to get the plots of the desired KPIs.
This master thesis explores a way to do surrogate modelling of the current process as is highlighted in Figure 1.6 by making use of GNN or MLP for the modelling of electrical engine designs described parameterically.

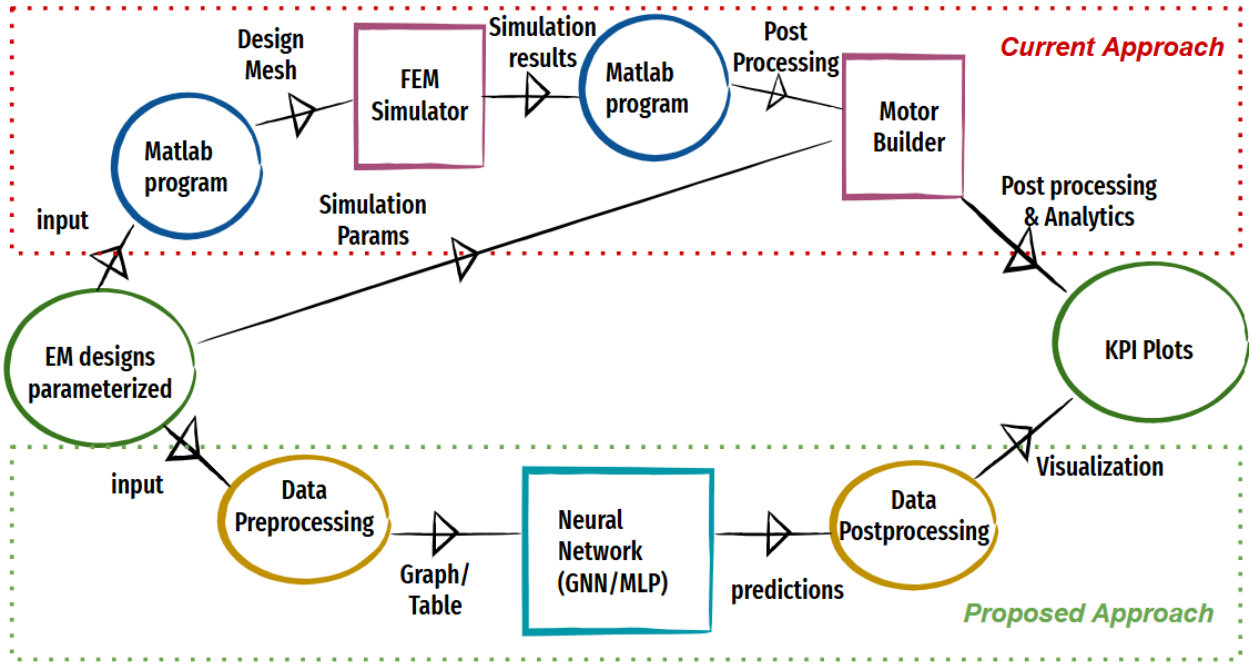



Figure 1.6: EM Design Flowchart

## 1.1 Objective

Our task is to predict 2 KPIs namely the Torque curve and the ETA grid from the parameteric description of topology invariant EM.
The remaining KPI such as Losses can be calculated from these 2 KPIs as they are inversely proportional

to the Efficiency values.
The Torque curve is a 2D plot ie, a vector across speed ranges and the Efficiency grid is a 3D plot ie, a matrix of dimensions torque ranges times speed ranges.
Both the outputs are continuous values which makes it a regression problem.

## 1.2 Problem Statement

1. We have a regression problem but as we are predicting two values it makes it a multi regression problem. The model architectures in Section 5.1 shines light on how we go about it.
2. Another concern we came across is the Torque Curve typically harbours integers values however our task is a regression problem and we discuss how we tackle it in Section 5.1.
3. The Efficiency grid dimensions vary across EM variants, and we need the target from the model to be of a fixed size. This problem is mitigated in Section 3.1.3
4. The ranges among the 1st and 2nd target vary significantly and is yet another challenge we overcome in Section 5.2.2
5. Furthermore, we presume graph representation of the data will be more logical than tabular representation due to its ability to grasp the connections within the EM structural properties better. This would be even more realistic with achieving topology invariance than the tabular representation and conserve memory and compute in the long run. However, we realize that our problem cannot be solved using Homogeneous GNN which is relatively simpler and is built on a single node and edge type. In order to model our problem as a graph, we need to represent it as a Heterogeneous graph.

## 1.3 Research Question

GNNs are relatively more expressive models as they can acapture the variability and complexity of our data better.
However, GNN in general have not been to less explored even so more the heterogeneous GNN. Particularly in the scenario of EM Modelling, there has been no publications with GNNs. Hence the need to check its feasibility and its performance with our benchmarks on tabular data. Additionally, existing Heterogeneous GNNs works e.g on recommendation networks, academic networks, information networks, social networks etc involves one large graph with multiple node and edge types. However, our problem involves creation of multiple heterogeneous graphs ie, 1 per EM variant.
Therefore, the applicability of Heterogeneous GNNs for our problem is to be seen.

We were skeptical on building a model architecture wherein the 2nd KPI is dependent on the 1st KPI as typically the former's shape is dependendent on the latter and not its values. However we can assume the values within the shape to be non zero values and otherwise 0s and model it in future.

Alternatively we could also build 2 models one for each KPIs and thus feed in the dependent predictions when training the latter.
However, it would be computationally expensive and does not help in the scenario when we might need to generate EM parameteric descriptions. Additionally we deemed it unnecessary as the dimensions of the ETA grid vary with the torque curve and not necessarily the ETA values.

Our thesis is loosely inspired from [05] which has strived to extend Message Passing (MP) into Heterogeneous GNNs.

## 1.4 Thesis Structure

Over the course of the thesis we shall refer the Torque curve as Mgrenz KPI and the Efficiency grid as ETA KPI respectively.
The remainder of the thesis is organized to follow sections namely Literature Review, Dataset, Graph Modelling, Modelling, Experiments and Results, Conclusion, and Bibliography.
In Literature Review section will introduce the works that has already been carried out in this domain.
In the Dataset section a detailed insight to how our data is structured is elaborated.
Graph Modelling section is an empirical study which outlines the background of GNNs primarily Heterogeneous GNNs and defines its concepts.
In the Modelling section, we introduce the network architectures and loss regularization techniques used to tackle the problem.
The outcomes of our work are presented in Experiments and Results chapter in addition to other findings we unearth.
Conclusion chapter summarizes the thesis briefly and would also give a glimpse into areas of improvement.
Finally the Bibliography section lists out the articles cited for this thesis.

# Chapter 2

# Literature Review

There has been extensive research in modeling the Electric Motor with Convolution Neural Network (CNN) based on the images of the motor cross-section.
However our approach is progressive in the sense that once the KPIs are predicted we would like to be able to generate the inputs.
Reproducing images is not known to be the best approach given the infamous known fact that AI generated images are faulty. However by generating the parameters of the motor we can be rest assured of more precise results.
Hence the need to focus on the inputs as they are with their parametric description.
Existing literature also covers works on modelling this work as tabular data using MLPs. Although this is fairly good forseeing the impact of generating the inverse process yet MLPs cannot necessarily learn all the intricacies within motor components.
Hence the need to better represent the data typically in the form of graphs and model GNNs to achieve the desired results.
There has been close to no work of GNNs in this domain. However we see progress of GNNs in molecular chemistry and social networks usecases from which we draw inspiration.

# Chapter 3

# Dataset

Valeo an automotive company has supplied the dataset consisting of close to 1500 Double V Electric Motor parameters. Around 89 parameters which comprises of the geometric, physical and simulation properties of the motor are chosen among the 196 parameters depending on its overall variability and significance. This was a design decision we made based on our understanding of the data.
Figure 3.1 shows the geometry of a whole Double V motor which can be sliced into 8 symmetrical parts.

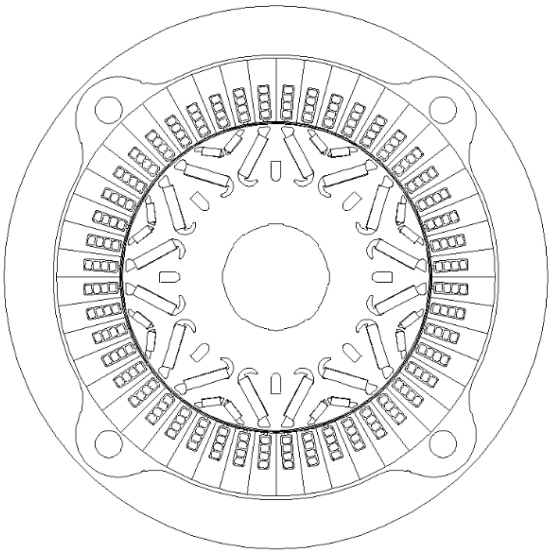


Figure 3.1: Complete EM Geometry(Source:Valeo)

Figure 3.2 shows our understanding of how the geometry of 1/8 cross-section of the same motor looks like. This comes in handy when creating the graph representation of the motor.

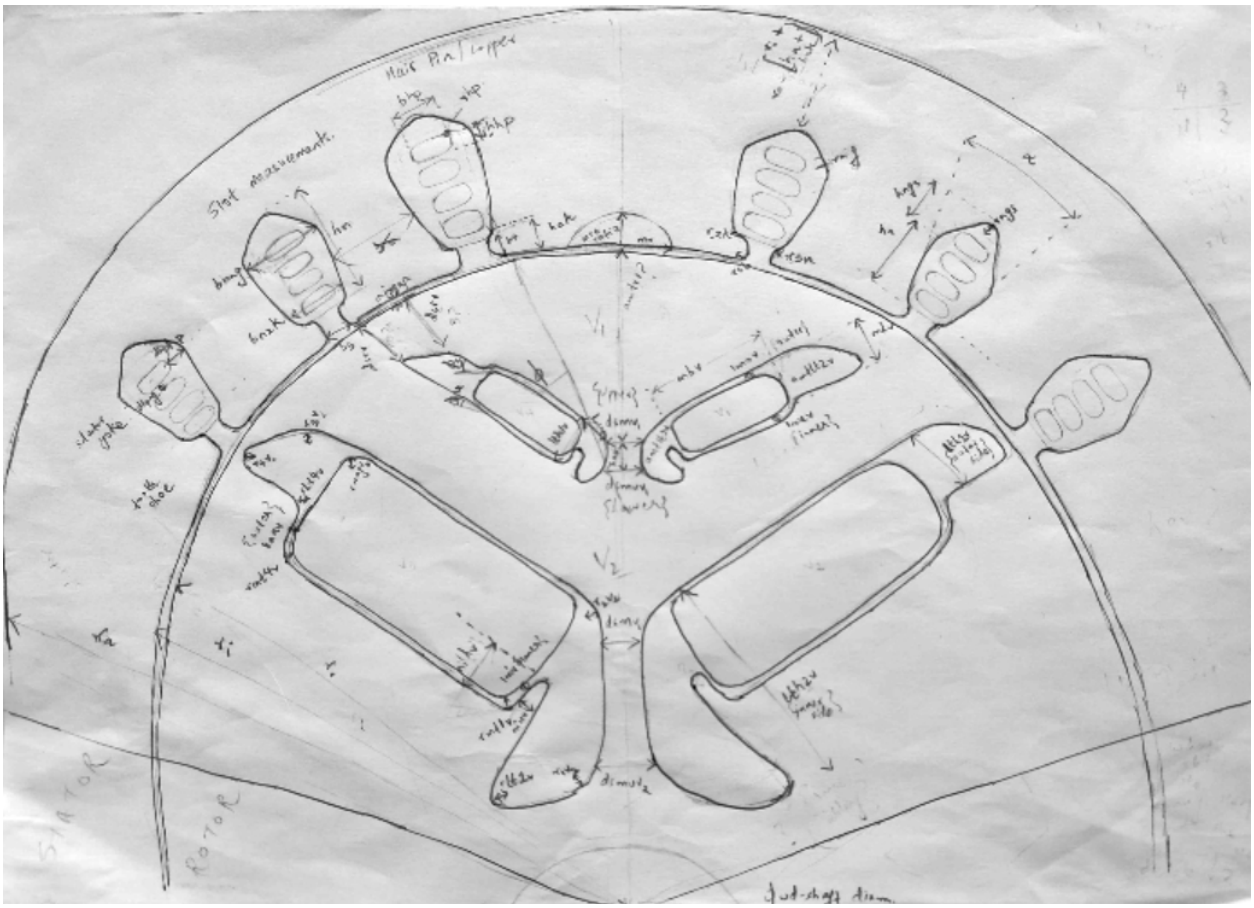



Figure 3.2: 1/8 Motor Crossection

Valeo has shared 1481 Excel Workbook files for each motor variant. Each of the excel files contain multiple sheets. Table 3.1 summarizes the sheets of interest to us:

| Sheet | Sheet Name | Description |
|---|---|---|
| Motor Parameters | input_data | Contains the input parameters for the motor model.<br>Includes geometric, physical, and simulation properties.<br>Unit dimensions, if applicable, are generally in mm or degrees. |
| Speed Grid | NN | Contains the speed grid.<br>Used for plotting the Mgrenz KPI and ETA KPI.<br>Unit dimensions is rpm. |
| Torque Grid | MM | Contains the torque grid.<br>Used for plotting the ETA KPI.<br>Unit dimensions is Nm. |
| ETA Grid | ETA | Contains the ETA KPI.<br>Has the same dimensions as the NN and MM sheets.<br>Percentage values do not have unit dimensions. |
| Torque Curve | Mgrenz | Contains the values corresponding to the Mgrenz KPI.<br>Has the same columns as the speed grid.<br>Unit dimensions is Nm. |

Table 3.1: Excel File Structure of an EM variant

## 3.1 Data Preprocessing for MLP

For modelling the MLP, we present the data in tabular form with the parameters corresponding to columns. In order to make the data compatible with our model, some level of data processing was carried out as elaborated below.

### 3.1.1 Data Exploration of the Input Parameters

All parameters including the additional ones in each topology are considered as a separate columns and therefore if a particular column is topology dependent the data of the other topologies for that corresponding column is treated as 0 values.
The values are read and stored in their float equivalent to preserve data precision.
Furthermore all degree columns are converted to their equivalent radian values as the latter are directly related to the geometry and makes derivation of other parameters much simpler whereas degrees is a notion of denoting a sliced angle.

Table 3.2 shows the summary statistics of the input parameters of the motor.

| Parameter | Unit of Dimension | Mean | Standard Deviation | Value Range | Single V Magnet Topology | Double V Magnet Topology | Nabla Magnet Topology |
|---|---|---|---|---|---|---|---|
| **General Parameters** | | | | | | | |
| N | # | 4.0 | 0.0 | 4.0–4.0 | ✓ | ✓ | ✓ |
| simQ | # | 6.0 | 0.0 | 6.0–6.0 | ✓ | ✓ | ✓ |
| r_a | mm | 9.000000e-02 | 2.554375e-15 | 9.000000e-02–9.000000e-02 | ✓ | ✓ | ✓ |
| r_i | mm | 0.064433 | 0.000902 | 0.064000–0.067000 | ✓ | ✓ | ✓ |

Continued on next page

**Table 3.2 – continued from previous page**

| Parameter | Unit of Dimension | Mean | Standard Deviation | Value Range | Single V Magnet Topology | Double V Magnet Topology | Nabla Magnet Topology |
|---|---|---|---|---|---|---|---|
| **Rotor Parameters** | | | | | | | |
| rad_phiv2 | rad | -0.511319 | 0.063678 | -0.645772–0.000000 | × | ✓ | × |
| lmsov2 | mm | -0.000289 | 0.000035 | -0.000300–0.000000 | × | ✓ | × |
| lth1v2 | mm | 0.005395 | 0.000359 | 0.000000–0.005450 | × | ✓ | × |
| lth2v2 | mm | 0.002789 | 0.000178 | 0.000000–0.002800 | × | ✓ | × |
| r1v2 | mm | 0.002097 | 0.000322 | 0.000000–0.002200 | × | ✓ | × |
| r11v2 | mm | 0.000326 | 0.000040 | 0.000000–0.000600 | × | ✓ | × |
| r2v2 | mm | 0.001873 | 0.000133 | 0.000000–0.001900 | × | ✓ | × |
| r3v2 | mm | 0.000697 | 0.000044 | 0.000000–0.000700 | × | ✓ | × |
| r4v2 | mm | 0.000747 | 0.000048 | 0.000000–0.000750 | × | ✓ | × |
| rmt1v2 | mm | 0.000249 | 0.000016 | 0.000000–0.000250 | × | ✓ | × |
| rmt4v2 | mm | 0.000249 | 0.000016 | 0.000000–0.000250 | × | ✓ | × |
| rlt1v2 | mm | 0.000185 | 0.000045 | 0.000000–0.000200 | × | ✓ | × |
| rlt4v2 | mm | 0.000249 | 0.000016 | 0.000000–0.000250 | × | ✓ | × |
| hav2 | mm | 0.004942 | 0.000336 | 0.000000–0.005000 | × | ✓ | × |
| mbv2 | mm | 0.017706 | 0.001175 | 0.000000–0.018100 | × | ✓ | × |
| mhv2 | mm | 0.003649 | 0.000265 | 0.000000–0.003800 | × | ✓ | × |
| rmagv2 | mm | 0.000498 | 0.000032 | 0.000000–0.000500 | × | ✓ | × |
| dsmv2 | mm | 0.002925 | 0.000194 | 0.000000–0.003100 | × | ✓ | × |
| dsmuv2 | mm | 0.002925 | 0.000194 | 0.000000–0.003100 | × | ✓ | × |
| amtrv2 | mm | 0.015888 | 0.001024 | 0.000000–0.016000 | × | ✓ | × |
| dsrv2 | mm | 0.000996 | 0.000064 | 0.000000–0.001000 | × | ✓ | × |
| lmav2 | mm | 0.00010 | 0.00003 | 0.00000–0.00011 | × | ✓ | × |
| lmiv2 | mm | 0.000109 | 0.000008 | 0.000000–0.000110 | × | ✓ | × |
| lmov2 | mm | 0.000055 | 0.000015 | 0.000000–0.000100 | × | ✓ | × |
| lmuv2 | mm | 0.000145 | 10.000017 | 0.000000–0.000150 | × | ✓ | × |
| rad_phiv1 | rad | -0.694266 | 0.054700 | -0.785398—0.453786 | ✓ | ✓ | ✓ |
| lmsov1 | mm | -0.000501 | 0.000094 | -0.000530–0.000500 | ✓ | ✓ | ✓ |
| lth1v1 | mm | 0.002889 | 0.000178 | 0.002855–0.005450 | ✓ | ✓ | ✓ |
| lth2v1 | mm | 0.002104 | 0.000059 | 0.002100–0.003200 | ✓ | ✓ | ✓ |
| r1v1 | mm | 0.000407 | 0.000108 | 0.000400–0.002200 | ✓ | ✓ | ✓ |
| r11v1 | mm | 0.000219 | 0.000045 | 0.000100–0.000600 | ✓ | ✓ | ✓ |
| r2v1 | mm | 0.000216 | 0.000100 | 0.000200–0.001900 | ✓ | ✓ | ✓ |
| r3v1 | mm | 0.000899 | 0.000013 | 0.000700–0.000900 | ✓ | ✓ | ✓ |
| r4v1 | mm | 0.000501 | 0.000016 | 0.000500–0.000750 | ✓ | ✓ | ✓ |
| rmt1v1 | mm | 2.500000e-04 | 8.839232e-18 | 2.500000e-04–2.500000e-04 | ✓ | ✓ | ✓ |
| rmt4v1 | mm | 2.500000e-04 | 8.839232e-18 | 2.500000e-04–2.500000e-04 | ✓ | ✓ | ✓ |
| rlt1v1 | mm | 0.000117 | 0.000056 | 0.000050–0.000250 | ✓ | ✓ | ✓ |
| rlt4v1 | mm | 2.500000e-04 | 8.839232e-18 | 2.500000e-04–2.500000e-04 | ✓ | ✓ | ✓ |
| hav1 | mm | 0.002918 | 0.000136 | 0.002900–0.005000 | ✓ | ✓ | ✓ |
| mbv1 | mm | 0.007643 | 0.000588 | 0.007500–0.018150 | ✓ | ✓ | ✓ |
| mhv1 | mm | 0.002808 | 0.000140 | 0.002700–0.005000 | ✓ | ✓ | ✓ |

Continued on next page

**Table 3.2 – continued from previous page**

| Parameter | Unit of Dimension | Mean | Standard Deviation | Value Range | Single V Magnet Topology | Double V Magnet Topology | Nabla Magnet Topology |
|---|---|---|---|---|---|---|---|
| rmagv1 | mm | 5.000000e-04 | 1.767846e-17 | 5.000000e-04–5.000000e-04 | ✓ | ✓ | ✓ |
| dsmv1 | mm | 0.001079 | 0.000141 | 0.000800–0.002800 | ✓ | ✓ | ✓ |
| dsmuv1 | mm | 0.001079 | 0.000147 | 0.000800–0.002920 | ✓ | ✓ | ✓ |
| amtrv1 | mm | 0.005538 | 0.000634 | 0.005500–0.019000 | ✓ | ✓ | ✓ |
| dsrv1 | mm | 0.000752 | 0.000032 | 0.000750–0.001250 | ✓ | ✓ | ✓ |
| lmav1 | mm | 0.000092 | 0.000026 | 0.000010–0.000110 | ✓ | ✓ | ✓ |
| lmiv1 | mm | 1.000405e-04 | 6.354235e-07 | 1.000000e-04–1.100000e-04 | ✓ | ✓ | ✓ |
| lmov1 | mm | 0.000055 | 0.000015 | 0.000050–0.000100 | ✓ | ✓ | ✓ |
| lmuv1 | mm | 0.000145 | 0.000015 | 0.000100–0.000150 | ✓ | ✓ | ✓ |
| rad_phi3b1 | rad | -0.002522 | 0.056001 | -1.256637–0.000000 | × | × | ✓ |
| rad_phi4b1 | rad | -0.000530 | 0.011775 | -0.261799–0.000000 | × | × | ✓ |
| lmsob1 | mm | 0.000002 | 0.000034 | 0.000000–0.000750 | × | × | ✓ |
| lthb1 | mm | 0.000006 | 0.000128 | 0.000000–0.002900 | × | × | ✓ |
| r2b1 | mm | 0.000002 | 0.000045 | 0.000000–0.001000 | × | × | ✓ |
| r3b1 | mm | 0.000002 | 0.000034 | 0.000000–0.001000 | × | × | ✓ |
| r4b1 | mm | 5.064146e-07 | 1.124422e-05 | 0.000000e+00–2.500000e-04 | × | × | ✓ |
| r5b1 | mm | 5.064146e-07 | 1.124422e-05 | 0.000000e+00–2.500000e-04 | × | × | ✓ |
| lgr3b1 | mm | 0.000001 | 0.000022 | 0.000000–0.000500 | × | × | ✓ |
| lgr4b1 | mm | 6.076975e-07 | 1.349307e-05 | 0.000000e+00–3.000000e-04 | × | × | ✓ |
| mbb1 | mm | 0.000030 | 0.000675 | 0.000000–0.015000 | × | × | ✓ |
| mhb1 | mm | 0.000006 | 0.000144 | 0.000000–0.003200 | × | × | ✓ |
| mtbb1 | mm | 0.000030 | 0.000675 | 0.000000–0.015000 | × | × | ✓ |
| rmagb1 | mm | 0.000001 | 0.000022 | 0.000000–0.000500 | × | × | ✓ |
| amtrb1 | mm | 0.000004 | 0.000098 | 0.000000–0.002500 | × | × | ✓ |
| dsr3b1 | mm | 0.000003 | 0.000066 | 0.000000–0.001850 | × | × | ✓ |
| dsr4b1 | mm | 0.000004 | 0.000083 | 0.000000–0.001850 | × | × | ✓ |
| lmob1 | mm | 2.025658e-07 | 4.497689e-06 | 0.000000e+00–1.000000e-04 | × | × | ✓ |
| lmub1 | mm | 3.038488e-07 | 6.746534e-06 | 0.000000e+00–1.500000e-04 | × | × | ✓ |
| lmsub1 | mm | 0.000004 | 0.000081 | 0.000000–0.001800 | × | × | ✓ |
| **Stator Parameters** | | | | | | | |
| airgap | mm | 1.000000e-03 | 3.535693e-17 | 1.000000e-03–1.000000e-03 | ✓ | ✓ | ✓ |
| b_nng | mm | 0.005049 | 0.000091 | 0.004646–0.005120 | ✓ | ✓ | ✓ |
| b_nzk | mm | 0.004557 | 0.000057 | 0.004450–0.004646 | ✓ | ✓ | ✓ |
| b_s | mm | 0.001002 | 0.000025 | 0.001000–0.001400 | ✓ | ✓ | ✓ |
| h_n | mm | 0.011149 | 0.000806 | 0.009200–0.013939 | ✓ | ✓ | ✓ |
| h_s | mm | 1.000000e-03 | 3.535693e-17 | 1.000000e-03–1.000000e-03 | ✓ | ✓ | ✓ |

Continued on next page

**Table 3.2 – continued from previous page**

| Parameter | Unit of Dimension | Mean | Standard Deviation | Value Range | Single V Magnet Topology | Double V Magnet Topology | Nabla Magnet Topology |
|---|---|---|---|---|---|---|---|
| r_sn | mm | 2.500000e-04 | 8.839232e-18 | 2.500000e-04–2.500000e-04 | ✓ | ✓ | ✓ |
| r_zk | mm | 0.000501 | 0.000019 | 0.000500–0.000800 | ✓ | ✓ | ✓ |
| r_ng | mm | 0.000501 | 0.000019 | 0.000500–0.000800 | ✓ | ✓ | ✓ |
| h_zk | mm | 1.000000e-03 | 3.535693e-17 | 1.000000e-03–1.000000e-03 | ✓ | ✓ | ✓ |
| bhp | mm | 0.003736 | 0.000081 | 0.003550–0.003800 | ✓ | ✓ | ✓ |
| hhp | mm | 0.002386 | 0.000199 | 0.001900–0.002840 | ✓ | ✓ | ✓ |
| rhp | mm | 0.000636 | 0.000045 | 0.000500–0.000800 | ✓ | ✓ | ✓ |
| dhphp | mm | 0.000264 | 0.000014 | 0.000263–0.000485 | ✓ | ✓ | ✓ |
| dhpng | mm | 0.000406 | 0.000003 | 0.000406–0.000453 | ✓ | ✓ | ✓ |

Table 3.2: EM Input Parameters

It gives us a clear idea of how the parameters of the Rotor can vary across Topologies. In addition the N and simQ parameters are count of Stator poles and its windings respectively and may vary across EM variants.
1481 examples are of the Double V Magnet Topology apart from which 3 examples each for the other 2 topologies. Hence, the imbalanced topologies parameters would have relatively not so reliable statistics.

### 3.1.2 Data Exploration of the Mgrenz KPI(Torque Curve)

Figure 3.3 shows the standard deviation of few samples of the Mgrenz KPI.

We make the below observations from it :

1. The Root Mean Squared Error (RMSE) is at its peak at low speeds.
2. The curve to an extent resembles a mirrored S shape. This finding is critical for how we modelled the loss regularization for the target and will be further elaborated in Section 5.2.1.

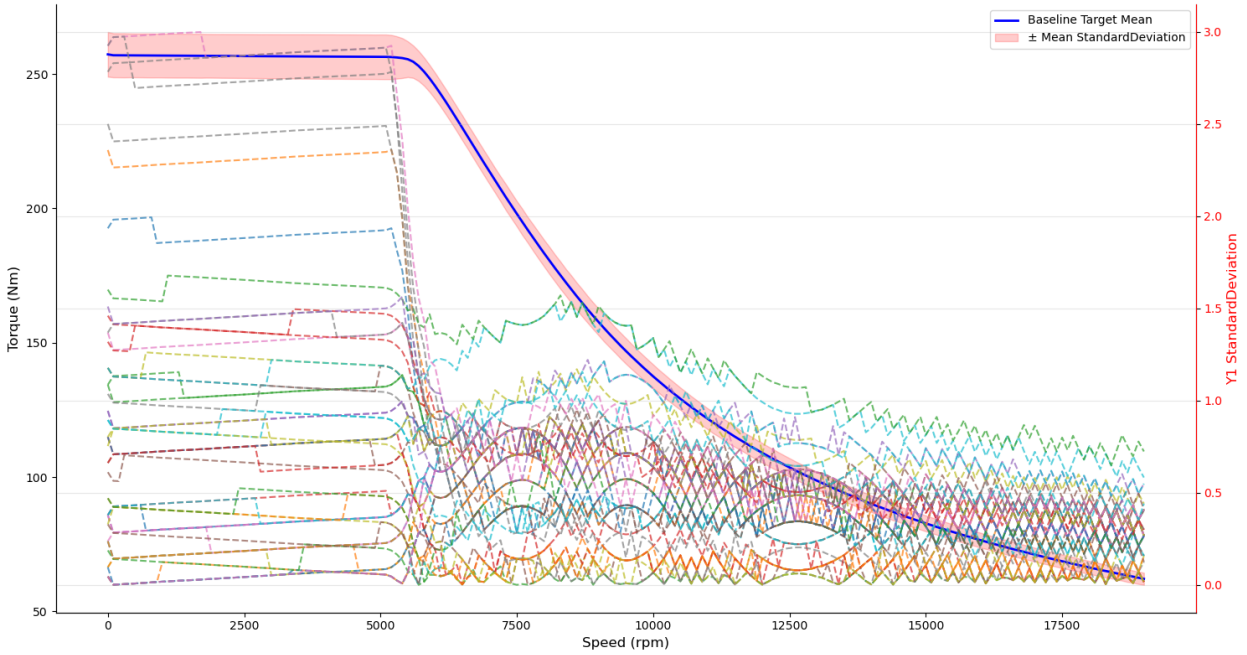



Figure 3.3: Standard Deviation of 2D KPI(ETA) Targets

### 3.1.3 Data Exploration of the ETA KPI(Efficiency Grid)

As the target values Mgrenz KPI and ETA KPI are not provided with the correct dimensions we have an additional step which takes the maximum torque value from the Mgrenz KPI and create a similar grid ranging from -maximum torque to maximum torque.
We then choose only the rows corresponding to this range from the actual MM grid supplied and the same row indices is used to retrieve the ETA KPI.
This step ensures that we grant the model the correct dimensions of the ETA KPI based on Torque KPI and predict likewise.

Efficiency values for negative torque values correspond to when motor is in generating mode and those of positive torque values when motor is in monitoring mode.
In both modes, the efficiency is almost similar however from Figure 3.4, we note it is not the case for our data. This is evident from low speed-high torque distribution area where we can see Not a Number (NaN) values.
Since these are FEM simulations, it is probably an effect of a post processing step taken by the Motor builder. This observation made us decide on dropping the negative ETA KPI and only predict the positive ETA KPI.
The latter can be mirrored in order to obtain the negative ETA KPI values if necessary. It can be mirrored to replicate the efficiency when it is in generating mode.

From Figure 3.5, we can observe the deviation is at its peak at low torques, low speeds and the border of the curve within the grid. We discuss the modelling of this information in Section 5.2.2.

Additionally the ETA KPI envelope is completely dependent on its equivalent Mgrenz KPI curve. The area beneath the boundary of which is looked into by the EM manufacturers to determine the car's efficiency in the operating cycle. This is yet another finding we use in Post Processing as is further elaborated in Section 5.5

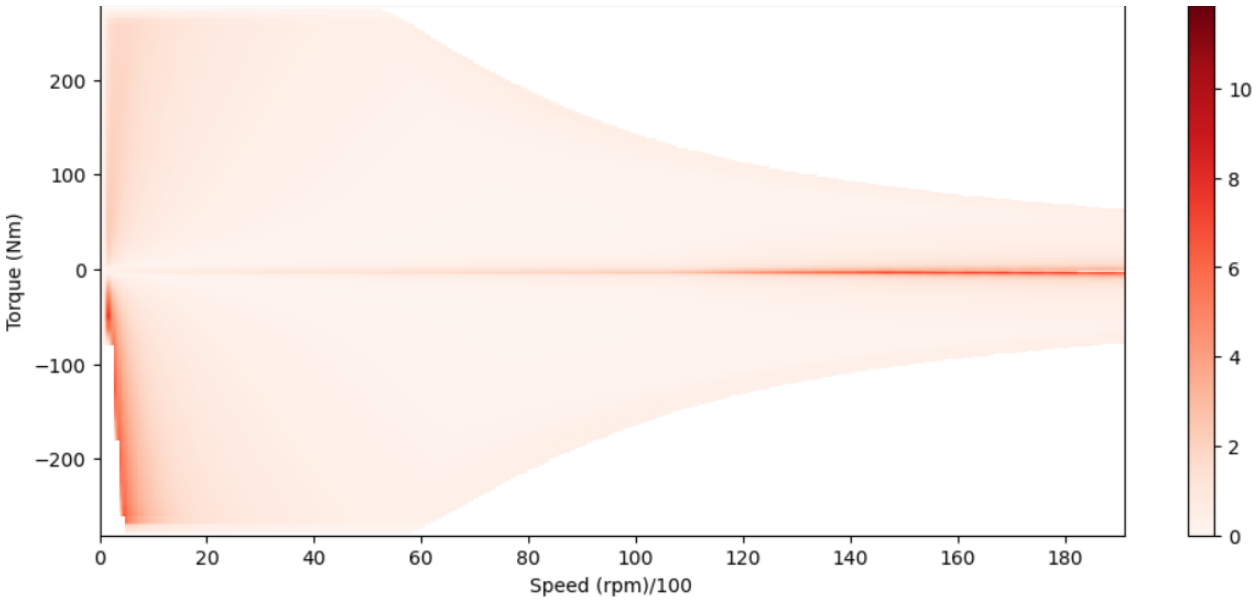



Figure 3.4: Standard Deviation of ETA KPI

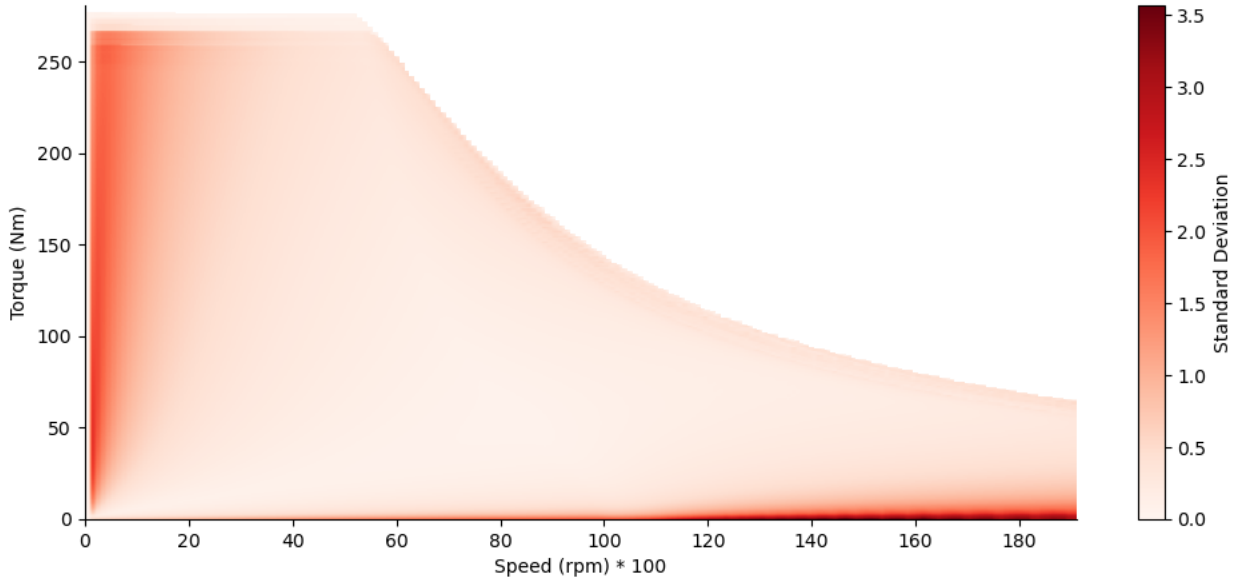



Figure 3.5: Standard Deviation of ETA KPI Positive Grid

Figure 3.6 gives us the big picture of how the efficiency values are distributed across equally spaced intervals of speed. The distributions shows skewness towards extreme speeds, extreme torques. We discuss in Section 5.2.2 how we integrate this finding into teaching over model. In addition, we notice skewness at the border of the curve within the grid, we discuss in Section 5.5 how this is learned.

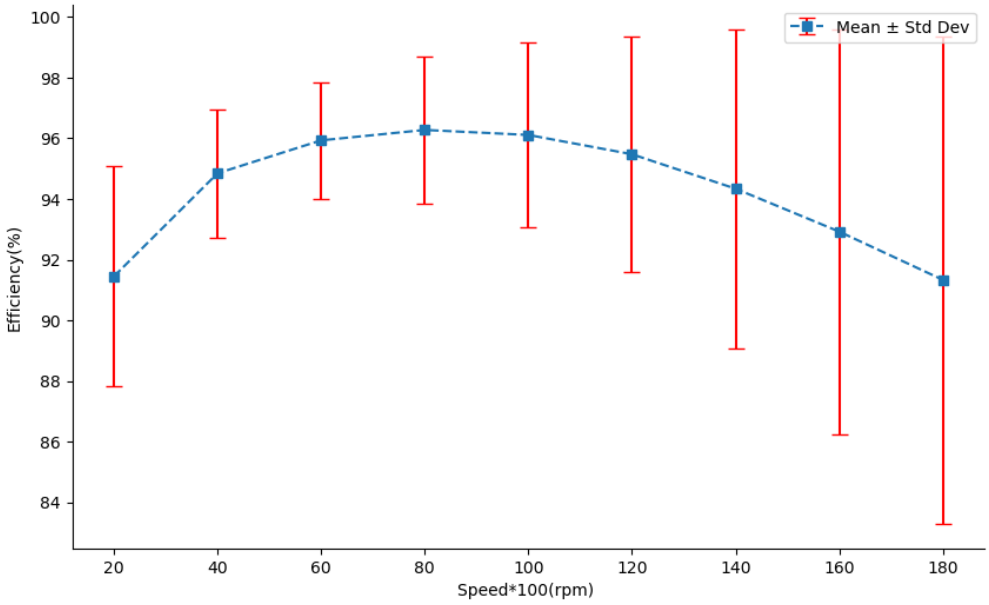



Figure 3.6: Standard Deviation of ETA KPI Positive Grid across Speed Intervals

Due to the infamous fact that reading sheets in excel files take up a lot of time and compute, we read the files as a onetime job when creating the table and store them into pythonic objects for faster access for training.
Both the input and target values for the Mgrenz KPI is stored locally as csv files whereas those of the ETA KPI is stored as separate csv files per variant considering it is in the form of a 2 Dimension (2D) array.
The csv files are then concatenated and stored into an array conserving dimensionality by padding NaN values to match dimensionality of the grid corresponding to the Mgrenz KPI with the largest torque value. In our case the value is 280 but this is subject to change as we receive more data and can be overridden by the user. The array is then saved locally for easy access and loading during training.

Initially we tried to set it NaN values as an incredibly high value hoping the model would consider this as a default value instead of NaN. However, it resulted in poor predictions as the model must have been confused and tried to increase its spread of predictions to cover this large value and so all true values were also predicted to be close to this dummy value. Fortunately, we have come up with a better way of handling this scenario and we discuss in Section 5.2.2 how we tackle it.

## 3.2 Scaling

Scaling is a common practice done before training a neural network. Standard scaling is the most prevalent scaling mechanism used for normalization as it results in a Gaussian Distribution centered around the mean. We have used the same for the input features to bring them to a common scale.

The Scaling is formulated mathematically as in Equation 3.1

$$z = \frac{x - \mu}{\sigma} \tag{3.1}$$

*where*

- $x$ : *Input*
- $\mu$ : *Mean*
- $\sigma$ : *Standard Deviation*

For the Input features both Mean and Standard deviation are calculated across columns. This is attributed to the fact we have columns with different ranges for the input.

We decided against scaling the targets owing to below 2 reasons :

1. They do not enter the network architecture but are only used during loss calculation.
2. If we scale the target we will have to scale each example from the train dataset and then average the calculated scaling parameters. It is not a good practice to do so as we will have lost a lot of originality in each example and is now introducing noise to new examples especially when they have a different data distribution.

During our experimentations where we initially scaled the targets, we observed that then the network required substantially less effort to learn and consequently lower learning rate and fewer epochs. Since this comes at a tradeoff of losing precision, we continued with the original targets.

## 3.3 Dataset splitting

We convert the data to float tensors for better precision and collate them into a Tensor Dataset. We have also partitioned the dataset to have about 50 samples for test and the remaining is used for 5 fold cross validation with 80:20 split for training and validation.
The reason we have a separate test dataset from the validation is to ensure that there is no data leakage as we do not want to overfit the test dataset with the hyperparameters we choose during training.

Within 5 folds, we expect to cover most grounds on training and have good monitoring on the model's performance for each fold. We have also used Data loaders to split the dataset into batches that fits into our Graphics Processing Unit (GPU) memory.

# Chapter 4

# Graph Modelling

We intended to model our problem as a Graph and solve using a GNN. We presume this will be a more clever way of representing our usecase in comparision to tabular data.. The idea was to make use of the Graph dynamics in our usecase and aggregate features that are semantically similar.

The node features captures the information of the node's role in the network and likewise does the edge features. They also take into account their respective neighbourhoods.

GNNs aim to learn a representation vector for each node using the MP algorithm. MP is based on the graph structure and initial node features. For each node in the graph over each successive hop of its neighbourhood until it has covered the entire graph, the model updates the learned node representation recursively aggregating its neighbour node features.
At the end of the MP algorithm, each node in the graph would have a good understanding of the other nodes within the same graph. Therefore the final representation will be rich enough to have captured all the information within the graph and be used for downstream tasks namely :

GNNs have several applications where data are generated from non-Euclidean domains and are represented as graphs with complex relationships and interdependency between objects [18].
GNNs applications range from Recommendation Systems, Chemistry, Traffic, Academic Networks, Information Networks to Social Networks [26].
These networks are generally represented by a giant graph in which case graph batching, splitting are unique. However our task requires us to build 1 graph each for each EM variant.

GNN predictions can be categorized broadly into : [18]

1. Node Level Prediction - represents the node classification and regression
2. Edge Level Prediction - focuses on edge classification and link prediction
3. Graph Level Prediction - relate to graph prediction tasks.
   Incidently our task is a Graph Level prediction.

Graphs in general can have the following taxonomy: [18]

1. Directed Graph - A graph in which the edges have a direction.
2. Undirected Graph - A graph in which for edge there will be its respective reverse edge in the opposite direction and the Adjacency matrix is symmetric.

In addition to their ability to incorporate both entity features and network features into a single, simultaneously trained model, most GNNs scale linearly with the number of edges in the network, making them applicable to large networks. A huge benefit of GNNs in practical use cases is that they are inductive rather than transductive. While transductive model can only be used on the specific data that was present during training, inductive models can be applied to entirely new data without having to be retrained. [05]

## 4.1 Introduction

We largely adopt the commonly used notations from CITE and reformulate it slightly to meet our needs: we use bold uppercase characters to denote matrices and bold lowercase characters denote vectors

A graph is represented as $G = (V, E)$ where $V$ is the set of vertices or nodes , and $E$ is the set of edges. Let $v_i \in V$ denote a node and $e_{ij} = (v_i, v_j) \in E$ denote an edge pointing from $v_j$ to $v_i$. The neighborhood of a node $v$ is defined as $N(v) = \{u \in V \mid (v, u) \in E\}$. The adjacency matrix $A$ is a $n \times n$ matrix with $A_{ij} = 1$ if $e_{ij} \in E$ and $A_{ij} = 0$ if $e_{ij} \notin E$. A graph may have node attributes $X$, where $X \in \mathbb{R}^{n \times d}$ is a node feature matrix with $x_v \in \mathbb{R}^d$ representing the feature vector of a node $v$. Meanwhile, a graph may have edge attributes $X_e$, where $X_e \in \mathbb{R}^{m \times c}$ is an edge feature matrix with $x_{v,u} \in \mathbb{R}^c$ representing the feature vector of an edge $(v, u)$.

A graph is represented by $G = (V, E)$, where $V$ is the vertex set with $|V| = n$ and $E$ is the edge set. In this paper, we consider undirected graphs. Denote by $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ the adjacency matrix, which is nonnegative. Denote by $D = \text{diag}(d_1, d_2, \ldots, d_n)$ the degree matrix of $A$, where $d_i = \sum_j a_{ij}$ is the degree of vertex $i$. The graph Laplacian (Chung, 1997) is defined as $L := D - A$, and the two versions of normalized graph Laplacians are defined as

$$L_{\text{sym}} := D^{-\frac{1}{2}} L D^{-\frac{1}{2}}.$$

A GN block contains three "update" functions, $\phi$, and three "aggregation" functions, $\rho$, [19]

$$e'_k = \phi_e(e_k, v_{r_k}, v_{s_k}, u)$$
$$v'_i = \phi_v(\bar{e}'_i, v_i, u)$$
$$u' = \phi_u(\bar{e}', \bar{v}', u)$$
$$\bar{e}'_i = \rho_{e \to v}(E'_i)$$
$$\bar{e}' = \rho_{e \to u}(E')$$
$$\bar{v}' = \rho_{v \to u}(V')$$

A hallmark of GNNs is that they can efficiently learn the architecture parameters from the training data [06].

GNNs cannot be too deep due to the oversmoothing problem. This is because the MP algorithm would have already traversed over all hops and made all features indistinguishable from one another. Besides a deep GNN would have a large number of parameters and would be computationally expensive to train.
[22] addresses this concern with the Graph Convolutional Network. Studies also tells us that it can be mitigated by including pre activation residual connections in the network.

Researchers have broadly classified GNNs following 2 distinct paradigms:

1. Homogeneous GNN
   Homogeneous GNN are designed for graphs with a single type of nodes and edges. MP is done for neighbouring nodes and edges over hops until it learns a representation equivalent from its neighbours. Homogeneous GNN are typically build to capture the structural information within a graph.
   These models are effective when the node and edge heterogeneity is negligible [01].

2. Heterogeneous GNN Heterogeneous GNN are designed for graphs with different type of nodes and edges. With different types of nodes and edges, also imply difference in features and dimensionality. Therefore, each type of node and edge reveal unique semantic information [24].
   As a single function cannot cater to each type, hence different MP and node updating functions needs to be implemented for each edge and node type respectively.
   Therefore MP is conditioned on the node and edge type thus allowing the flow of information to be more controlled.
   In addition to the structural information, Heterogeneous GNNs also excel to capture semantic information within the graph.

## 4.2 Heterogeneous GNN

Graphs are ubiquitously heterogeneous because of its capability to have different node and edge types on top of its inherent graph feature to model relations between objects.
Heterogeneous GNNs also takes into account homophily ie., nodes close on a network have similar embeddings [26] this is in essence the structural property and is true for all GNNs.
A heterogeneous graph can be defined as $G = \{V, E, \phi, \psi\}$ [02], where:
$V$ is the set of nodes, with a node type mapping function $\phi : V \rightarrow T_v$,
$E$ is the set of edges, with an edge type mapping function $\psi : E \rightarrow T_e$.

Each node $v_i \in V$ is assigned a node type $c_i = \phi(v_i) \in T_v$. [02]
Each edge $e_{t \leftarrow s} \in E$ (denoted $e_{ts}$ for short) is assigned a relation $r_{c_t \leftarrow c_s} = \psi(e_{ts}) \in T_e$ (or $r_{c_t c_s}$ for short), pointing from the source node $s$ to the target node $t$. [02]

The sets of possible node types can be represented as: $T_v = \{\phi(v) : \forall v \in V\}$.
The sets of possible edge types can be represented as: $T_e = \{\psi(e) : \forall e \in E\}$.
When $|T_v| = |T_e| = 1$, the graph degenerates into a homogeneous graph [02].

GNNs aim to learn a representation vector $\mathbf{h}_v^{(L)} \in \mathbb{R}^{d_L}$ for each node $v$ after $L$-layer transformations, based on the graph structure and the initial node feature $\mathbf{h}_v^{(0)} \in \mathbb{R}^{d_0}$ [04].

Heterogeneous GNN generally work by having separate non linear functions convolve over each edge type during message computation and over each node type when aggregating the learned information.
The graph structure of $G$ can be represented by a series of adjacency matrices $\{A_r : r \in T_e\}$. For each relation $r_{c_t c_s} \in T_e$, $A_{c_t c_s} \in \mathbb{R}^{|V_{c_t}| \times |V_{c_s}|}$ is the corresponding adjacency matrix where the nonzero values indicate positions of edges $E_{c_t c_s}$ of the current relation.

A metapath defines a composite relation of several edge types, represented as $P \equiv c_1 \leftarrow c_2 \leftarrow \ldots \leftarrow c_l$ ($P = c_1 c_2 \ldots c_l$ for short). They are composite relationships between nodes that help to capture the structural information of heterogeneous graphs.

The Metapaths in Heterogeneous GNNs are divided into 2 streams:

1. Metapath based method
   It aggregates the neighbourhood features of the same semantic first and then fuses different semantic information.

2. Metapath free method
   It aggregates message from neighbourhood of the same hop for all node types and so captures both structural and semantic information simultaneously.

   Multiple MP operations are done across different metapaths which are finally aggregated into a new representation for the node.[05].

Each metapath captures the proximity of nodes in a graph from a specific semantic view [26].

## 4.3 GNN Literature Review

Existing works [21] on modelling heterogeneous graphs usually split the graph into multiple homogeneous subgraphs based on their node types each retaining an aspect of the heterogeneity. This is ineffective in exploiting hidden rich semantic associations between different types of edges for large scale multi relational graphs.
[20]'s work does not undertake this methodology and is also independent of metapaths making it effective in dealing with large number of complex relations.
Bias amplification can be handled much better when a graph is can be is modelled as a heterogeneous graph [03].
Modelling Heterogeneous Graphs with customized metapath generally require domain specific knowledge. In order to combat this, [23] has used meta relations instead and a distinct edge based matrix for each edge

type.
Works by [25] also use meta relations to extract soft metapaths and does message passing over multi hop neighborhood of nodes.
Metarelations are triples n a heterogeneous network and is most often used in knowledge graphs which contain rich schema. Such meta relations are shallow embeddings and not deep as the meta paths [26].

[24] proposes a heterogeneous network node classification model which converts the heterogeneous network into multiple semantic graphs through meta-paths. The polynomial graph convolutional kernel is also described in this paper.

Most of the existing works also do not consider the edge features in their model. However [05] uses edge features as transactions in a financial network to detect cases of money laundering. It also breaks down the large heterogeneous graph into subgraphs based on of different node edge type combinations. The paper also brings to light that GraphSAGE does not incorporate the edge weights.

Message Passing Neural Network (MPNN) framework, unifying a large group of different GNN models. Apart from the unifying framework, the most essential contribution of this model is in our view that it applies a learned message-passing function that utilizes edge features.

GCN achieves this by breaking down the heterogeneous graph into multiple homogeneous ones, one for each edge type. In each layer, GCN is applied to each homogeneous graph, and the resulting node embeddings are element-wise summed to form the final output. drawback of RGCN is that it does not take node heterogeneity into account.

Additionally [01] has used heterogeneous GNNs to model the heterogeneity and sparsity in Electronic Health Records of patients. This work takes into account the relational features of the dataset which is arguably the most important factor in this domain. Their work also claim to learn its multi-task nature. However we note that this work works on meta relations but donot factor in edge features.

[27] introduces modelling that makes it capable of learning features of molecular graphs directly and is invariant to graph isomorphism. They also suggest that edge features in the network can be learned by introducing hidden states for all edges in the graph and updating them. It also implements a message passing network which factors in the edge types.

A *heterogeneous graph* is defined as a directed graph $G = (V, E, A, R)$ where each node $v \in V$ and each edge $e \in E$ are associated with their type mapping functions $\tau(v) : V \to A$ and $\varphi(e) : E \to R$, respectively.
For an edge $e = (s, t)$ linked from source node $s$ to target node $t$, its *meta relation* is denoted as

$$\langle \tau(s), \varphi(e), \tau(t) \rangle.$$

Naturally, $\varphi(e)^{-1}$ represents the inverse of $\varphi(e)$.
The classical *meta path paradigm* [17–19] is defined as a sequence of such meta relations.

This is in essence the relation triple ie., $(nodetypeofs, edgetype, nodetypeoft)$
**Definition 1. Heterogeneous Networks**. Given a graph $G = (V, E, A, R)$, if the graph has a node type mapping function $\epsilon : V \to A$ and a link type mapping function $\xi : E \to R$, where $A$ and $R$ are predefined node type sets and link type sets, and $A = 2^{R^*}$, the network is called a *heterogeneous network*.
**Definition 2. Metapath**. Given a path

$$A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} A_{l+1}$$

(abbreviated as $A_1 \xrightarrow{R_1 \cdots R_l} A_{l+1}$), if the path describes the composite relationship $R_1 \circ R_2 \circ \cdots \circ R_l$ between nodes of type $A_1$ and $A_{l+1}$, then the path is called a *meta-path*.
**Definition 3. Metapath-based Neighbors** . Given a node $m_1$ and a meta-path $\epsilon$ in a heterogeneous graph, the neighbors $N^{\epsilon}_{m_1}$ of node $m_1$ based on the meta-path $\epsilon$ are defined as the set of nodes connected to

$m_1$ via the meta-path $\epsilon$. This set of neighbors based on the meta-path also includes $m_1$ itself.

In HAN a meta-path M, the representation of node u is aggregated from its meta-path based neighbors NM(u) = u union v—v connects with u via the meta-path M.
In RGCN, during message passing, neighbors under the same edge type will be aggregated and normalized first.
GraphSAGE does not utlize edge features.

[26]

## 4.4 EM Heterogeneous GNN Model

We find the Heterogeneous graph to be most apt for our use case with its different node and edge types as it preserves both the structural and semantics of our data.
This property is crucial in modelling our use case as we will then have similar node-edge types per topology. In contrast Homogeneous graphs would lead to suboptimal results as it cannot factor in the heterogeneity and thus the semantic nature of the usecase fully.

Given the EM data D, our goal is to construct a heterogeneous graph G from D. Let T1, T2 on G be the 2 KPIs we need to learn from D. We aim to train a multi-task GNN model M such that M can deliver high performance on T1 and T2 .

### 4.4.1 EM Heterogeneous Graph Construction

Inspired by the promising advantages of HGNN, we took the effort of modelling one for only the Double V Magnet Topology.
We construct the heterogeneous graph adaptable for each of the EM variants as a NetworkX graph.

1. Node Types
   $T_v = \{v, vm, r, s, sw\}$
   The node type mapping functions of each node type in the graph are as follows :
   (a) $T_1$ : Rotor Airgap ($v$)
   $\phi = \{v11, v12, v21, v22\}$
   (b) $T_2$ : Rotor Magnet ($vm$)
   $\phi = \{v1m1, v1m2, v2m1, v2m2\}$
   (c) $T_3$ : Radius ($r$)
   $\phi = \{rr, ra, o\}$
   (d) $T_4$ : Stator Poles ($s$)
   $\phi = \{s1, s2, s3, s4, s5, s6\}$
   (e) $T_5$ : Stator Windings ($sw$)
   $\phi = \{s1w1, s1w2, s1w3, s1w4,$
   $s2w1, s2w2, s2w3, s2w4,$
   $s3w1, s3w2, s3w3, s3w4,$
   $s4w1, s4w2, s4w3, s4w4,$
   $s5w1, s5w2, s5w3, s5w4,$
   $s6w1, s6w2, s6w3, s6w4\}$
2. Edge Types
   $T_e = \{a, d1, d2, d4\}$
   The edge type mapping functions of each edge type in the graph are as follows :
   We define the meta relations as well for each of the edge type

(a) $T_1$ : Angular/Radian Edges ($a$)

(b) $T_2$ : Distance Edge with 1 feature ($d1$)

(c) $T_3$ : Distance Edge with 2 feature ($d2$)

(d) $T_4$ : Distance Edge with 4 features ($d4$)

3. Node Features

4. Edge Features

Figure 4.1 shows the heterogeneous graph we constructed

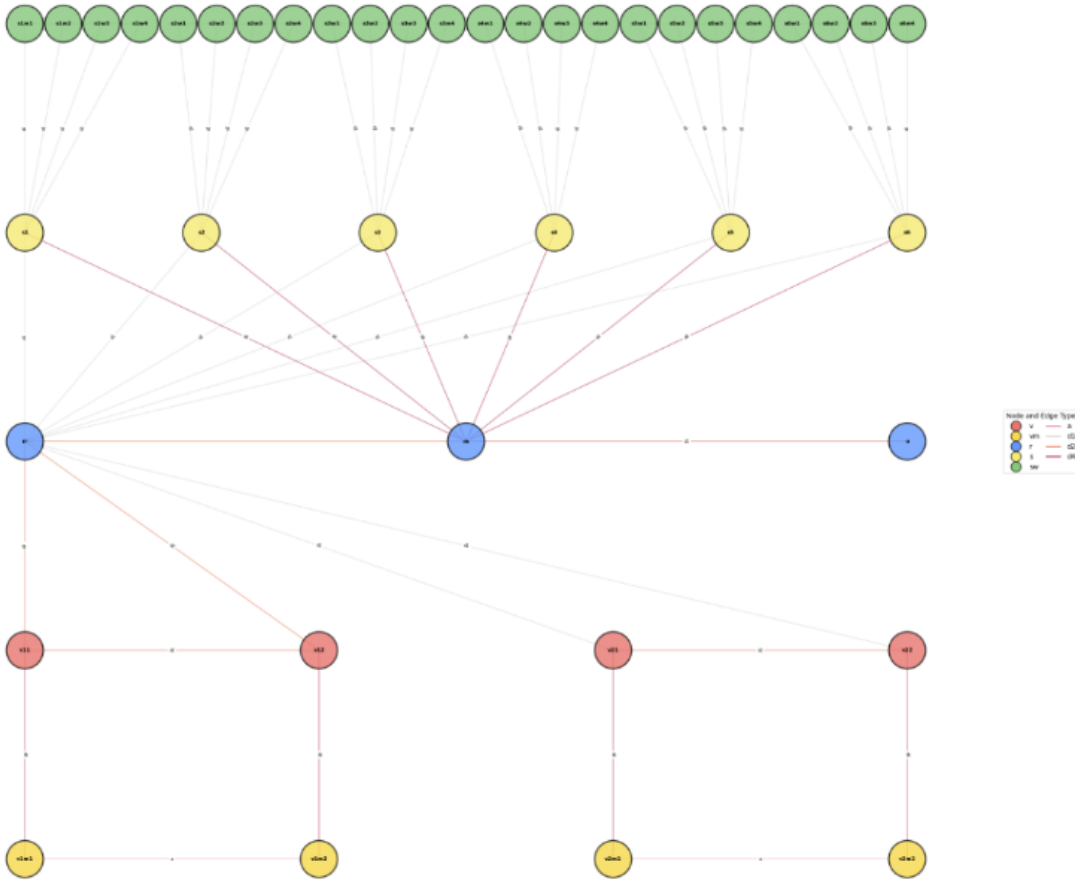


Figure 4.1: HetGraph

The count of certain parameters within the motor such as stator poles with its corresponding stator windings and rotor magnets is made more comprehendable to the model having new nodes and edges whereas for the MLP architecture this information is represented only as a number in yet another column.
By leveraging the relational features introduced by the meta relations, we can obtain a good graph representation of the data.

We then learn the constructed heterogeneous graph through a heterogeneous GNN model.

# Chapter 5

# Modelling & Evaluation

For our multi regression problem, we first train a MLP on the tabular representation of the data and work on it further to do the same with a heterogeneous GNN.

## 5.1 MLP Model

For the MLP model, we use a single model with input features corresponding to all the features in the tabular topology invariant representation of the data.
The model architecture is build to predict both the Mgrenz KPI and ETA KPIs by having 2 separate output layers for each of the KPIs.
Since the Mgrenz KPI's targets are relatively learnable than that of the ETA KPI's targets we have experimented with fewer feed forward layers in the former than in the latter.
We have a hyperparameter to control the number of neurons in each hidden layer this can be tuned and is further discussed in Section 6.1.
Rectified Linear Unit (ReLU) layers were also added in between to serve as the activation function and produce non-linearities and so noise in the network.
Dropout layers ensure that not all neurons in each layer are used up during training to prevent the model from memorizing the data and hence overfitting. We have 2 hyperparameters to control the dropout rate at which we freeze the neurons when training also to be discussed in Table 6.1. The 2 dropouts are for shared layers of the MLP and for the layers corresponding to ETA KPI.
Batch normalisation layers are used to normalize the input from the ReLU activations applied on it and so mitigate internal covariate shift to the next layer and hence speed up the training process.
Thus both batch normalisation and dropout layers stabilize the network training.

Figure 5.1 gives an outline on how the MLP Model architecture is designed.

We discuss each component of the architecture below :

1. Input
   The input layer takes in all features of the tabular data which is 89 in our case as scaled tensors.

2. MLP Shared
   The MLP Shared block is a sequential block comprising of 2 Linear Layers with the input features and neurons of each hidden layer to be a learnable size we tune. We donot increase the number of neurons in the hidden layers within this block as it needs to be in the range of input features and output features(in this case Mgrenz KPI). Furthermore we have Batch Normalisation layers between each linear and ReLU activation function in addition to drop out layers. The dropout rate for the layers in this block is relatively higher as we want to encourage the model to focus largely on learning the generality of the data. The 2 Linear Layers enable the network at the start to learn a rich represntation of the data at the intial feature extraction phase.

3. MLP Mgrenz
   This block comprises of Sequential 1 Linear Layer with the output feature to be the size of the Mgrenz KPI and a ReLU activation function. We use a ReLU activation function at the end of the output layer as the targets are inherently always positive values and it encourages the model to adhere to this fact.

4. MLP ETA
   This block comprises of Sequential 3 Linear Layer with the output feature to be the size of the ETA KPI. Here we also increase the neurons of the hidden layers as we are not limited by the dimensionality of the output feature. This would enable the model to be more strong and grasp the complex patterns in the data better. As usual we have batch normalisation, dropout and ReLU activation functions between the 1st two Linear layers. The dropout rate for the layers in this block is relatively lower as we want to encourage the model to learn the specific nature of the grid towards the end. For the Last Linear layer we have the output features corresponding to the target dimensions and a ReLU activation again as the targets are inherently always positive values and it again encourages the model to adhere to this fact.

5. Mgrenz KPI
   The number of output features correspond to the target size 191. Although the targets for the Mgrenz KPI are an array of integer values, we use the float tensor and not integer tensor to represent the data else it would become a classification problem and not a regression problem as it should be.

6. ETA KPI
   The number of output features correspond to the target size which in our case is the shape of the collated array we created as was discussed in Section 3.1.3

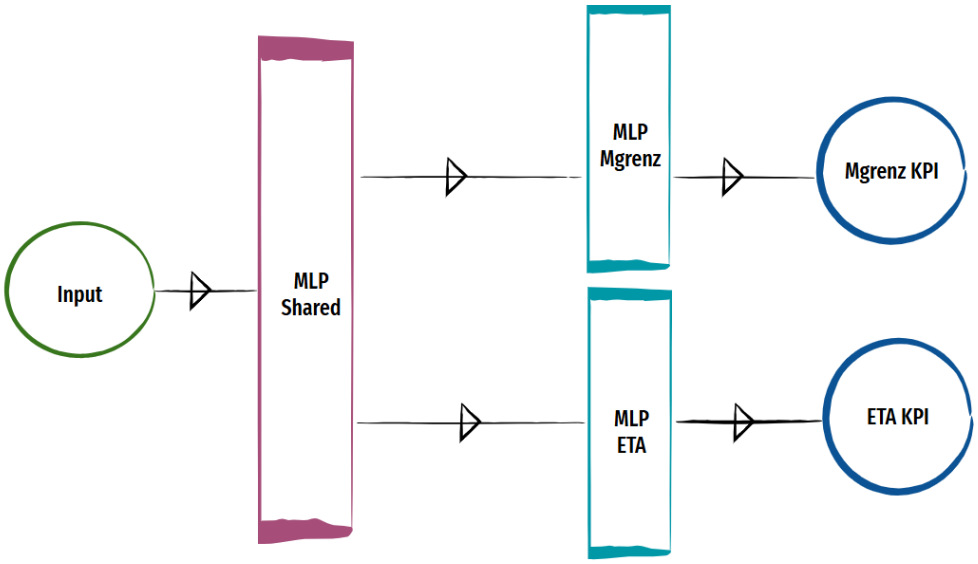



Figure 5.1: MLP Model Architecture

## 5.2 Loss Functions

The Mean Squared Error (MSE) loss is the loss function used for our problem with the intention that the squared losses penalize the model and inturn encourage it to minimize it further. In addition to its contribution in exaggerating the loss, MSE also ensures that deviations are positive and do not confuse the model by negating the losses of different signs.

### 5.2.1 Loss for Mgrenz KPI(Torque curve)

The MSE loss for the Mgrenz KPI is formulated mathematically as in Equation 5.1

$$\text{Y1 Loss} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}\sum_{j=1}^{h}(y_{ij} - \hat{y}_{ij})^2 \tag{5.1}$$

*where*

- n : EM samples
- h : Columns of 1D vector
- $\hat{y}$ : Prediction
- y : Target

To encourage the model to learn the nature of the curve, we have experimented with 2 Loss Regularization techniques. Both of which are L2 Regularization to be in sync with the dynamics of the MSE loss.

1. Smoothening Loss Regularization

   To smoothen out the curve for the Mgrenz KPI we apply the below loss regularisation factor to take into account.
   This is formulated mathematically as in Equation 5.2.

$$\text{Y1 Smoothening Loss Regularization} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h-1}\sum_{j=1}^{h-1}\left(\text{ReLU}(|\hat{y}_{ij+1} - \hat{y}_{ij}| - 1)\right)^2 \tag{5.2}$$

   This nature is factored in by penalising the loss by the magnitude if the neighbouring values in the prediction are not close to each other.
   ReLU helps to clips the difference if it is negative which is the scenario when a violation is not warranted.

2. Decreasing Loss Regularization

   The Torque curve closely resembles a decreasing sigmoidal curve and hence we use this knowledge to penalize the loss for non-decreasing values within each prediction.
   This is formulated mathematically as in Equation 5.3.

$$\text{Y1 Declining Loss Regularization} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h-1}\sum_{j=1}^{h-1}\left(\text{ReLU}(\hat{y}_{ij+1} - \hat{y}_{ij})\right)^2 \tag{5.3}$$

   It takes into consideration the almost continuous decreasing nature of the curve as the regularization is such that a specific element in the array is less than or equal to its prior element.

We have not combined both the above regularizations as they donot complement each other. This is because the loss regularized by Equation 5.2 will not necessarily be a decreasing curve. This holds true for the regularization in Equation 5.3 as it may not necessarily have gradual transitions in the curve. Nevertheless, we perform ablation studies with both the regularizations and report the results obtained in Table 6.3

### 5.2.2 Loss for ETA KPI(Efficiency Grid)

The MSE loss for the 3 Dimension (3D) KPI is formulated as in Equation 5.4.

$$\text{Y2 Loss} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{w}\frac{1}{h}\sum_{j=1}^{w}\sum_{k=1}^{h}(M_{ijk}\cdot y_{ijk}) - (M_{ijk}\cdot \hat{y}_{ijk}))^2 \tag{5.4}$$

$$M_{ijk} = \begin{cases} 1 & \text{if } y_{ijk} \neq \text{NaN} \\ 0 & \text{if } y_{ijk} = \text{NaN} \end{cases} \tag{5.5}$$

*where*

- $M_{ijk}$ *: Mask matrix*
- $w$ *: Rows of 2D vector*
- $h$ *: Columns of 2D vector*

The ETA KPI is a 3D plot of real numbers representing percentage values and is always in the range of 0-100%.
We noticed in some portions of the ETA KPI, the plot not visible as it had NaN values.
As ANN cannot be trained to predict NaN values we have a binary mask constructed such that values corresponding to NaN in the target have value 0 and all other values as 1. Mathematically, this process can be expressed as is in Equation 5.5 and thus ensure that the NaN values are ignored in the loss calculation. The mask is then multiplied with both the target and its respective prediction.

Additionally, to encourage the model to learn the nature of the ETA KPI from our observations gathered in Section 3.1.3, we have tried to incorporate all of the below learnings via the loss function as Y2 Regularization.

1. Efficiency at Maximum Torque Loss Regularization

   To ensure that the shape of the ETA KPI is maintained, we also regularize the loss for the maximum torque value. To do so, we have attempted to retrieve the last rows our ETA KPI and those of its target values and penalise the squared difference to have higher weight.

   We formulate it mathematically as in Equation 5.6

$$\text{Y2 Loss Regularization MM Max Mgrenz} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{t_1}\sum_{j=-t_1}^{w}\frac{1}{h}\sum_{k=1}^{h}(y_{ijk} - \hat{y}_{ijk})^2 \tag{5.6}$$

   *where*

   - $t_1$ *: Threshold for inital ETA KPI Envelope boundary*

   The number of last rows is determined by a threshold $t_1$

2. Efficiency at low Speeds

   To force the model to pay more attention at lower speeds, we have regularized the loss for the first few columns of each row of the ETA grid and penalise the squared difference with that of the target.

   We formulate it mathematically as in Equation 5.7:

$$\text{Y2 Loss Regularization Low Speed} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{w}\sum_{j=1}^{w}\frac{1}{t_2}\sum_{k=1}^{t_2}(y_{ijk} - \hat{y}_{ijk})^2 \tag{5.7}$$

*where*

- $t_2$ *: Threshold for Low Speed*

The number of first columns is determined by a threshold $t_2$.
It is a known fact that at 0 Torque, the corresponding efficiency values for the motor is 0. With this regularization, this learning as well is incorporated into the loss function.

3. Efficiency at low Torque

At extreme speeds we find a greater deviation in the efficiency values particularly towards higher speeds. This is because we have fewer efficiency values as speed increases beyond a range asnot all torque values participate. To force the model to be more careful at low torque, we have regularized the loss for the first few rows of each column of the ETA KPI and penalise the squared difference with that of the target.

We formulate it mathematically as in Equation 5.8:

$$\text{Y2 Loss Regularization Low Torque} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{t_3}\sum_{j=1}^{t_3}\frac{1}{h}\sum_{k=1}^{h}(y_{ijk}-\hat{y}_{ijk})^2 \tag{5.8}$$

*where*

- $t_3$ *: Threshold for High Torque*

The number of first rows is determined by a threshold $t_3$.

The above Y2 Regularizations are indeed purely MSE but with higher weights for specific regions of the ETA KPI. Consequently being L2 Regularizations it goes hand in hand with MSE Loss calculated in Equation 5.4.

We aggregate all the above regularizations to form the Y2 Loss Regularization as in Equation 5.9

$$\begin{aligned}\text{Y2 Loss Regularization} = \text{Y2 Loss Regularization MM Max Mgrenz}\\ +\text{Y2 Loss Regularization Low Speed} + \text{Y2 Loss Regularization Low Torque}\end{aligned} \tag{5.9}$$

The Total Loss is calculated in Equation 5.10

$$\begin{aligned}\text{Total Loss} = \text{wt} \times (\text{Y1 Loss} + (\lambda_{1y1} \times \text{Y1 Smoothening Loss Regularization}) + (\lambda_{2y1} \times\\ \text{Y1 Declining Loss Regularization })) + (\text{1-wt}) \times (\text{Y2 Loss} + (\lambda_{y2} \times \text{Y2 Loss Regularization}))\end{aligned} \tag{5.10}$$

*where*

- $\lambda_{1y1}$ : Y1 Smoothening Loss Regularization Parameter
- $\lambda_{2y1}$ : Y1 Declining Loss Regularization Parameter
- $\lambda_{y2}$ : Y2 Loss Regularization Parameter
- wt: Y1 Loss Weightage
- 1-wt: Y2 Loss Weightage

We have added a Weightage parameter that controls the contribution of the Y1 Loss and Y2 Loss to the Total Loss. There are 2 reasons why this is useful for us :

1. When the targets are not of the same scale.
   Without scaling, the losses for both targets being of different ranges are drastically different.
   We circumvent this by weighing up the loss of the target not performing better on validation dataset and weighing down the loss of the targets by the factor of how much its value range varies.

2. When the prediction accuracy of one KPI is substantially more important than the other.
   Our task demands the same as the ETA KPI is post processed to be within the shape of the Mgrenz KPI.
   Therefore, in theory have higher weightage for the Mgrenz KPI as its loss in performing well is costlier but as its value range is relatively higher we make decisions from monitoring the perdiction performances.
   We reflect on these decisions based on whether the envelope of the ETA KPI grid is more valuable that the efficiency values within it.
   To counter the Mgrenz KPI loss dominating the total loss, we have implemented quite a few regularization techniques to the ETA KPI loss.
   Ultimately we decided on prioritizing the efficiency values.

Furthermore, the weightage parameters for both target is designed to sum upto 1 keeping in mind improved training stability as a result of normalized weights.

Finally the aggregated loss is backpropagated.

## 5.3 Optimizer

Adam optimizer is used for optimization as it is known to be computationally efficient and requires little memory [17].
It infers the gradients of the loss and how it impacts the weights and biases of each layer and thus guide the model to decrease the loss.
The optimizer acts once the loss is backpropagated across training each batch of the dataset.
The optimizer also uses the learning rate to control the step size with which the model parameters are updated.

## 5.4 Evaluation Metrics

The evaluation metrics we have considered for our regression problem is the average of the RMSE.
Therefore, the model with the least prediction scores ie, closest to 0 is ideal for our application.

### 5.4.1 Evaluation Metrics for Mgrenz KPI

The Y1 Score for the Mgrenz KPI is formulated in Equation 5.11 :

$$\text{Y1 score} = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\sqrt{\frac{1}{h}\sum_{j=1}^{h}(y_{ij}-\hat{y}_{ij})^2}}_{Y1\ RMSE} \tag{5.11}$$

*where*

- *h : Columns of 1D vector*
- *Y1 RMSE: RMSE for each test sample*

### 5.4.2 Evaluation Metrics for ETA KPI

The Y2 score for the ETA KPI is formulated in Equation 5.12 :

$$\text{Y2 score} = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\sqrt{\frac{1}{w}\frac{1}{h}\sum_{j=1}^{w}\sum_{k=1}^{h}(y_{ijk}-\hat{y}_{ijk})^2}}_{Y2\ RMSE} \tag{5.12}$$

*where*

- $w$ *: Rows of 2D vector*
- $h$ *: Columns of 2D vector*
- *Y2 RMSE : RMSE for each test sample*

## 5.5 Post Processing

The mean and standard deviation from the train-validation datasets are applied to transform the test dataset to maintain uniformity in the predictions generated. In the case of new files we first convert it into the tabular representation our model consumes and then apply the scaling.
Hence the reason why we preserve the same scalers used during training as we not only evaluate our dedicated test dataset but also for clients to use on demand.

Furthermore as we are predicting a padded matrix to ensure dimensionality sync across different ETA KPI's, the grid contains values even outside the boundary of the ETA KPI. Hence, we attempted to slice the shape of the Mgrenz KPI curve from the ETA KPI by counting the number of columns a row to have based on consecutive values in the curve. This brings us back to the point that it is imperative the prediction of the Mgrenz KPI is close to perfect as the envelope of the ETA KPI inherently is dependent on it.

# Chapter 6

# Experiments and Results

## 6.1 Experiments with MLP

After an exhaustive random grid search of the hyperparamter space, we present the parameters in Table 6.1 by monitoring the model's performance across 5 fold cross validation training.
We use different splits for each round of training and tune the hyperparameters on the validation set. The final splits are saved locally and can be used later to ensure reproducibility.

| **Hyperparameters** | **Description** | **Value** | **Value Ranges** |
|---|---|---|---|
| lr | Learning Rate | 0.075 | 0.025 - 0.25 |
| hidden size | Dimensionality of Hidden Layers | 128 | 64, 128 |
| lr gamma | Exponential Learning Rate Scheduler Gamma Parameter | 0.9 | 0.5-0.9 |
| batch size | Batch Size | 72 | 32-72 |
| epochs | Number of Epochs | 10 | 6-10 |
| $p_{y1}$ | Dropout Probability for Shared Layers | 0.35 | 0.4-0.2 |
| $p_{y2}$ | Dropout Probability for ETA Layers | 0.2 | 0.2-0.1 |
| $\lambda_{1y1}$ | Y1 Smoothening Curve Loss Regularizer | 0.5 | 0-0.75 |
| $\lambda_{2y1}$ | Y1 Decreasing Curve Loss Regularizer | 0.5 | 0-0.75 |
| $\lambda_{y2}$ | Y2 Loss Regularizer | 3.75 | 0-5 |
| $t_1$ | Y2 Initial Envelope Boundary Threshold | 5 | 0-5 |
| $t_2$ | Y2 Low Speed Threshold | 20 | 0-30 |
| $t_3$ | Y2 Low Torque Threshold | 20 | 0-30 |
| wt | Weightage of Y1 Loss | 0.05 | 0-1 |
| 1-wt | Weightage of Y2 Loss | 0.95 | 0-1 |

Table 6.1: Hyperparameter Tuning

We are using an exponential learning rate scheduler which reduces the learning rate exponentially by the *lr gamma parameter* to decay learning as training progresses across epochs. This is to ensure that the model does not overshoot after few cycles of training.
The *batch size* is limited to the capacity of our GPU memory. We use the maximum batch size to train faster and therefore use 72 after which we hit the memory roof of our GPU.
The *hidden size* refers to the number of neurons in the hidden layers of the MLP model. We have experimented with 2 hidden sizes only as we are constrained with the fact that the number of neurons must be between the range of input features and output features which was discussed in Section 5.1. In addition to it being of the multiples of 8 as GPUs are most optimized for the same. Dropout rate during training is controled by the parameter $p_{y1}$ for shared layers and $p_{y2}$ for layers specific to ETA KPI. This is discussed in Section 5.1.
We donot concern much about the data being dropped in larger extent than anticipated for the Mgrenz

KPI because it is relatively simpler to learn and the value range is comparitively larger leading its loss to dominate the total loss.
$\lambda_{1y1}$, $\lambda_{2y1}$ and $\lambda_{y2}$ parameters control the regularization weight for the regularization terms for the Mgrenz KPI and ETA KPI respectively detailed in Equation 5.10.
The thresholds $t_1$, $t_2$ and $t_3$ are used to control the number of rows and/or columns to be considered for the regularization terms for the ETA KPI as detailed in Section 5.2.2.
The weightage parameter to control the direction of loss is denoted by $wt$ and is also discussed in Equation 5.10.

We also observe that the simple architecture of the model is not a significant downside for that limited data set.

We chose `Wandb`[1] to log metrics from the training run and to monitor model performance across the 5 folds.
Our MLP model only uses the loss regularization parameters $\lambda_{2y1}$ and $\lambda_{y2}$ ie, it does not consider the Smoothening curve Regularization fot the Mgrenz KPIs.

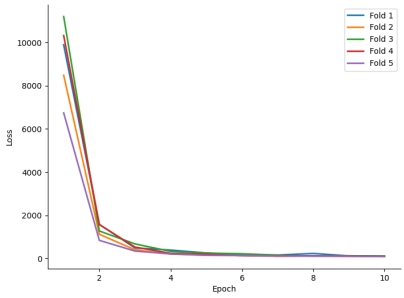



Figure 6.1: Aggregated Training Loss

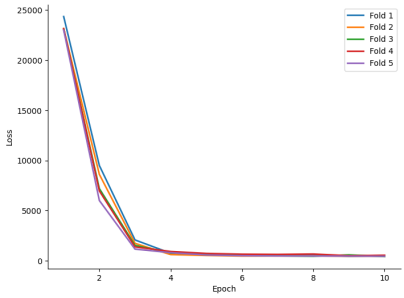



Figure 6.2: Training Loss for Mgrenz KPI

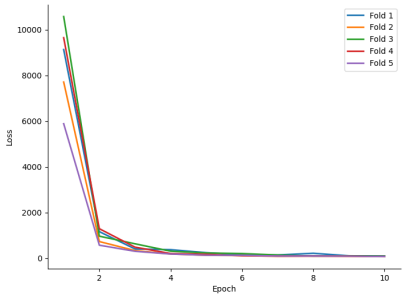



Figure 6.3: Training Loss for ETA KPI

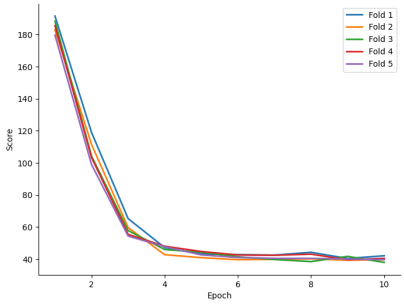



Figure 6.4: Aggregated Training Score

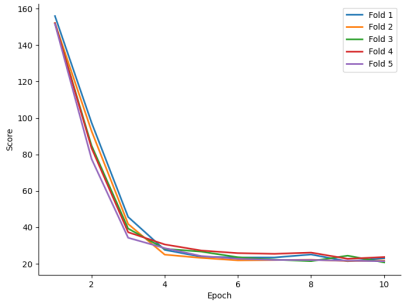



Figure 6.5: Training Score for Mgrenz KPI

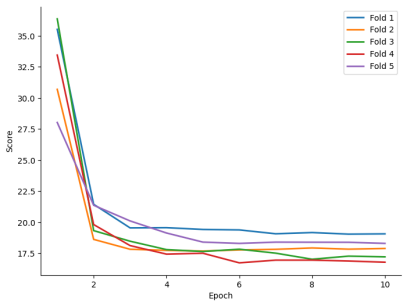



Figure 6.6: Training Score for ETA KPI

From the training plots we see that the model has converged after having run for 10 epochs.
Even though the weightage assigned to the ETA KPI is significantly more in addition to loss regularization parameter, it seems to struggle to learn beyond a threshold and overfit by the 5th epoch.
We hypothesize to further increase the weightage of the ETA KPI. The Mgrenz KPI shows promise in learning better but it would be at the cost of overfitting the ETA KPI.
The Validations plots tells us the same tale although the ETA KPI stops to learn after a certain point.

[1] Weights & Biases

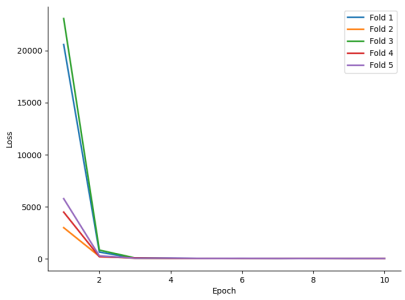



Figure 6.7: Aggregated Validation Loss

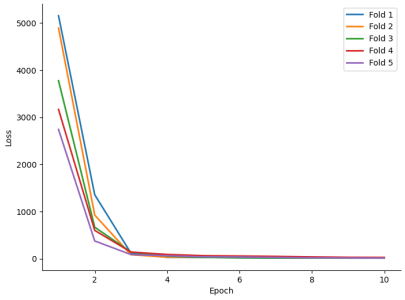



Figure 6.8: Validation Loss for Mgrenz KPI

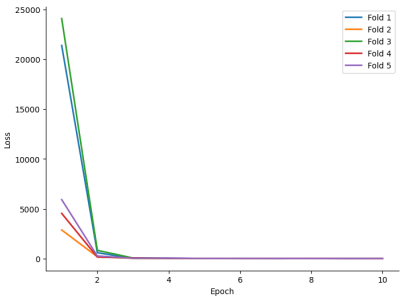



Figure 6.9: Validation Loss for ETA KPI

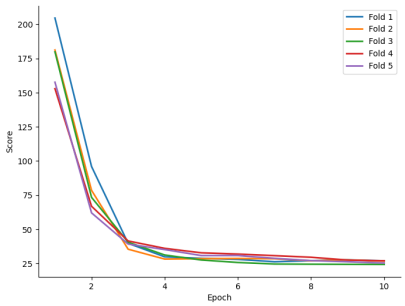



Figure 6.10: Aggregated Validation Score

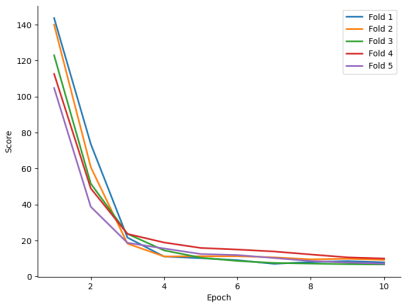



Figure 6.11: Validation Score for Mgrenz KPI

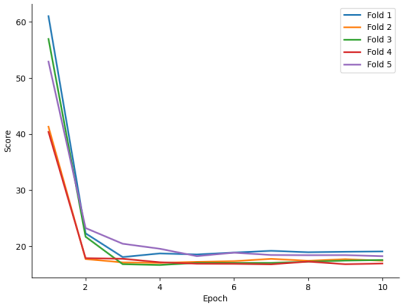



Figure 6.12: Validation Score for ETA KPI

We have also enabled saving the best performant fold locally so it can be loaded on demand by the client when in need to only run inference.
We have narrowed down scoring to follow the criteria as recorded in Table 6.2.

| **% Difference** | **0-5%** | **5-10%** | **10-15%** | **15-20%** | **20-25%** | **25-30%** | **30-35%** | **35-40%** | **40-100%** |
|---|---|---|---|---|---|---|---|---|---|
| Y1 Score | 0-11 | 11-22 | 22-33 | 33-44 | 44-55 | 55-66 | 66-77 | 77-88 | >88 |
| Y2 Score | 0-5 | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 | >40 |

Table 6.2: Scoring Criteria

This is deduced from the Equation 6.1

$$\text{Percentage Difference} = (\text{Score}/(Max - Min)) \times 100 \tag{6.1}$$

From our observations the target values for Mgrenz KPI range between 50-280 and those of the ETA KPI range between 0-100.

## 6.2 Results with MLP

### 6.2.1 Mgrenz KPI Results with MLP

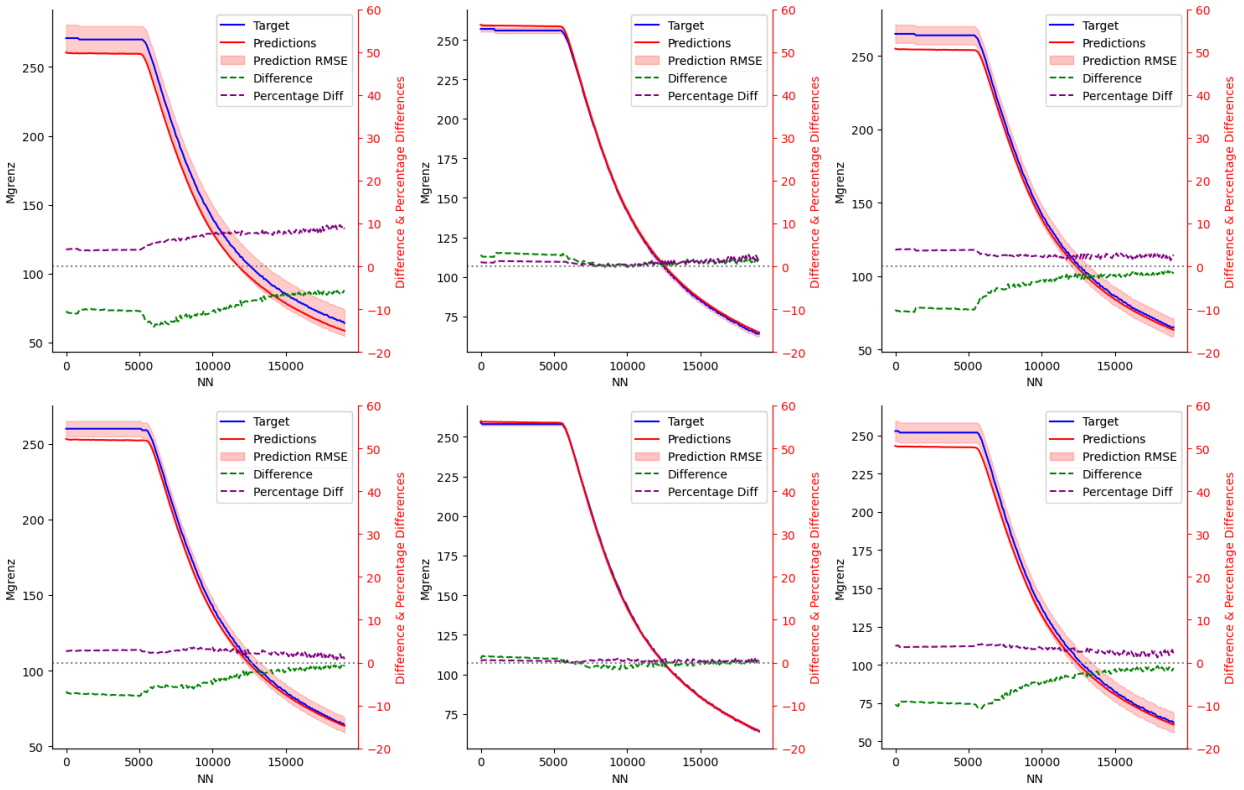



Figure 6.13: MLP Training Results for Mgrenz KPI

Figure 6.13 depicts the difference and percentage difference on a twin scale to give a rough overview of the prediction deviations from the targets. Whereas the RMSE equivalent to the Y1 score tells us that for 50% of the plots shown, the predictions are off by about 5% from the target values as per Table 6.2.
Figure 6.14 shows us the Average RMSE and element wise RMSE for the test dataset performance with the MLP.
Our inferences are overall the predictions closely resemble the trajectory of the target values allthough they fluctuate. Experimenting with the hyperparameters $\lambda_{1y1}$ and $\lambda_{2y1}$ has potential to improve this anomaly. In addition granting a higher weight $wt$ can also help in this direction.

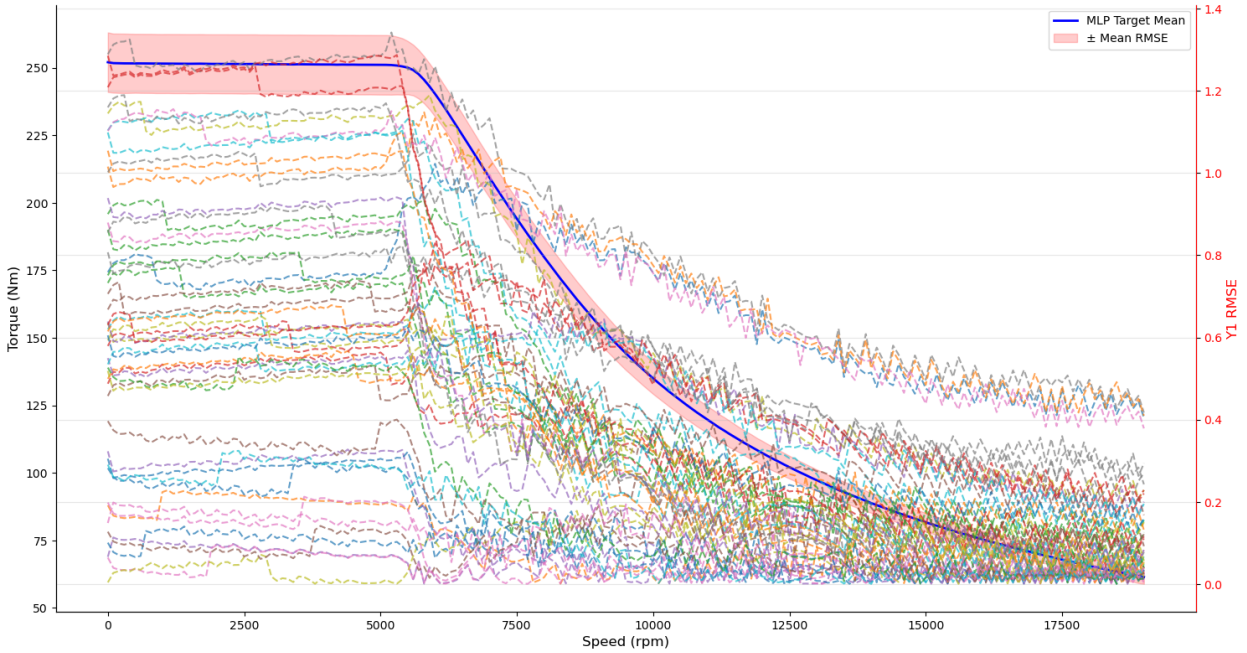



Figure 6.14: MLP RMSE Evaluation for Mgrenz KPI

Figure 6.15 shows the score statistics of the model performance of Mgrenz KPI over the entire test dataset. The Y1 RMSE from Equation 5.11 is calculated for each sample and shown as a histogram.

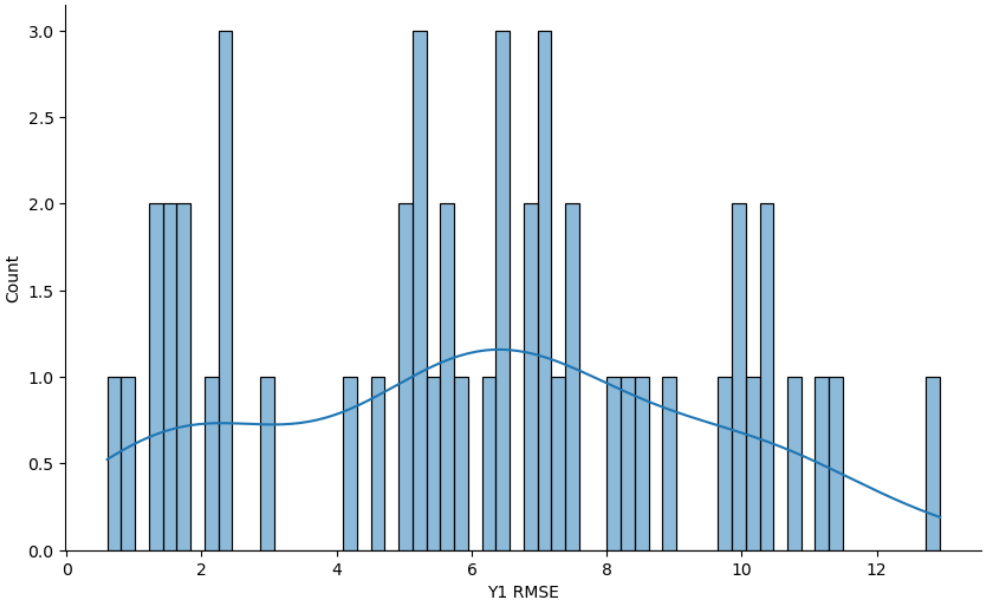



Figure 6.15: MLP Score statistics for Mgrenz KPI

It also tells us that the RMSEs are almost evenly distributed between 0.5-12.5 thus encompassing a range of 0-6% difference with the target values.

### 6.2.2 ETA KPI Results with MLP

The results of the MLP model from inference with its corresponding overlap is shown below:

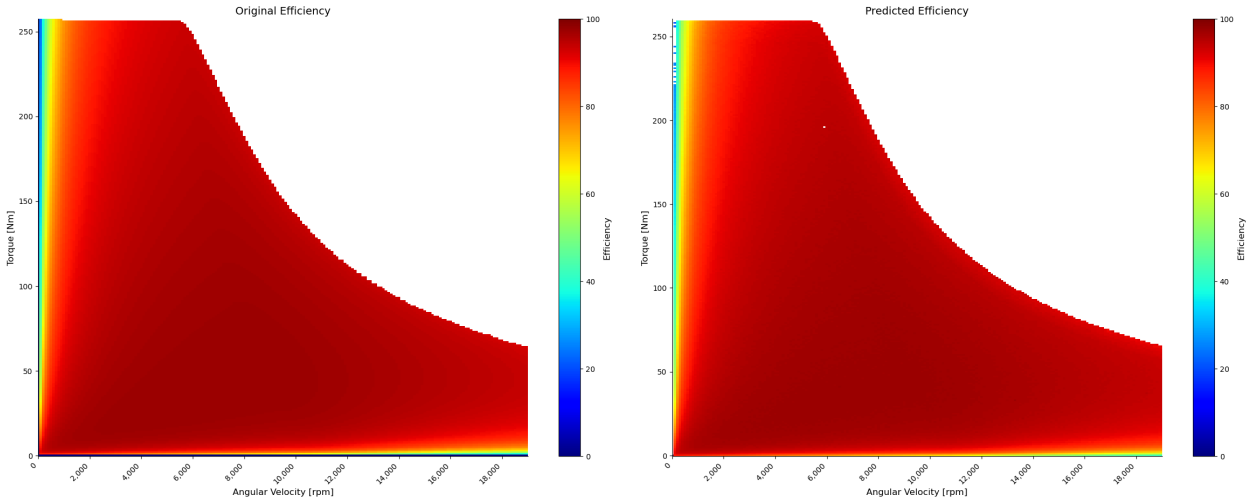



Figure 6.16: 1st MLP Training Results for ETA KPI

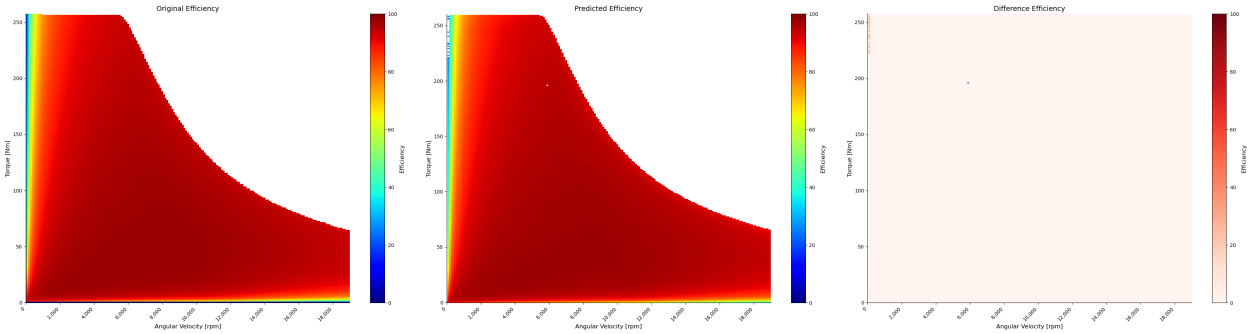



Figure 6.17: 1st Eval MLP Training Results for ETA KPI

From a visual perspective, the predictions to relative extent resembles the target and the regularization of the ETA KPI seems to be already teaching the model that the envelope should follow the Mgrenz KPI curve shape. The curve being sliced off irregularly is the effect of us trying to retain the Mgrenz KPI shape as was discussed in Section 5.5.
This could be having negative impacts for the ETA KPI when the predictions for the Mgrenz KPI are not perfect. This is yet again a decision to be made based on which of the 2 targets to prioritize.

Moreover we calculate the RMSE and differences of prediction from the target by truncating the matrices to be of common shape. We also replace NaN values with 0 in either if it occurs as the ETA KPI's were padded with NaN's to be of the same shape but originally had no values as was discussed in Section 3.1.3.

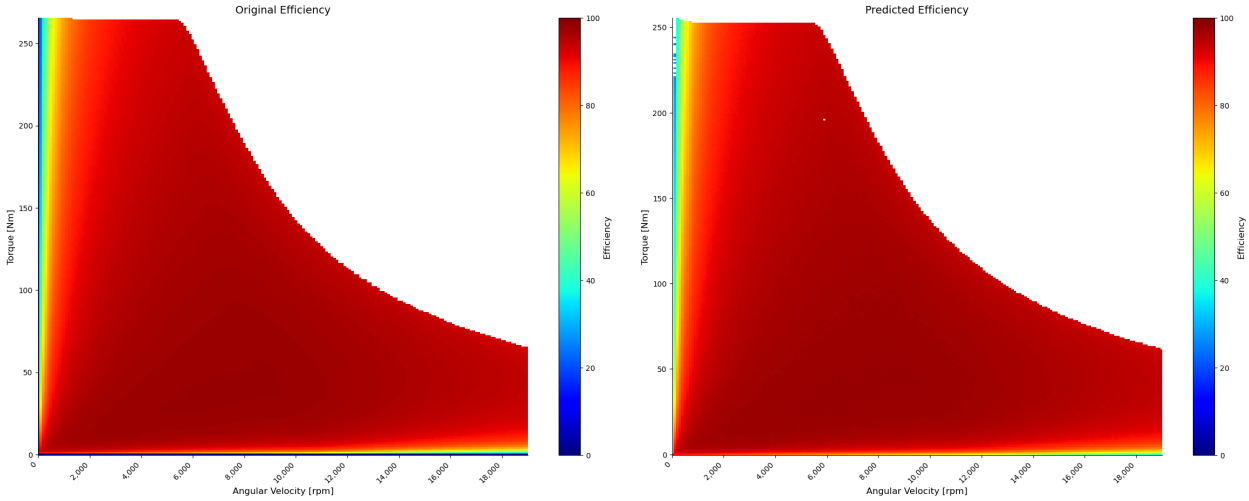



Figure 6.18: 2nd MLP Training Results for ETA KPI

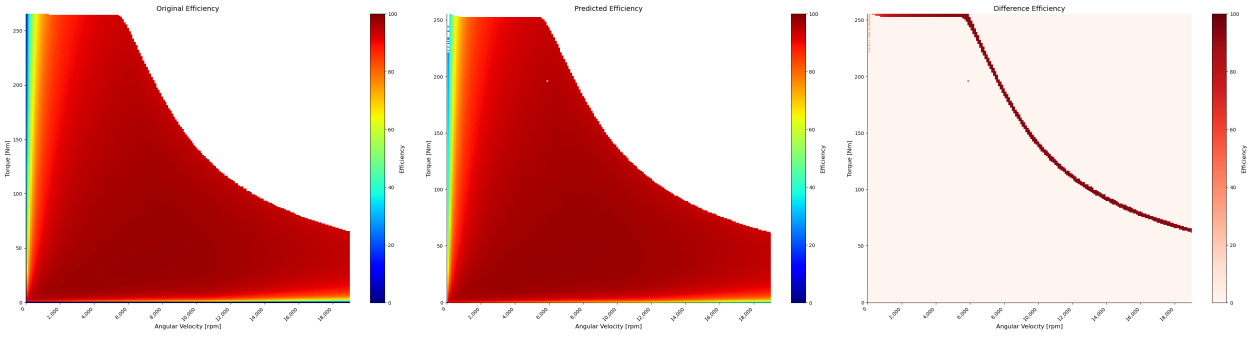



Figure 6.19: 2nd Eval MLP Training Results for ETA KPI

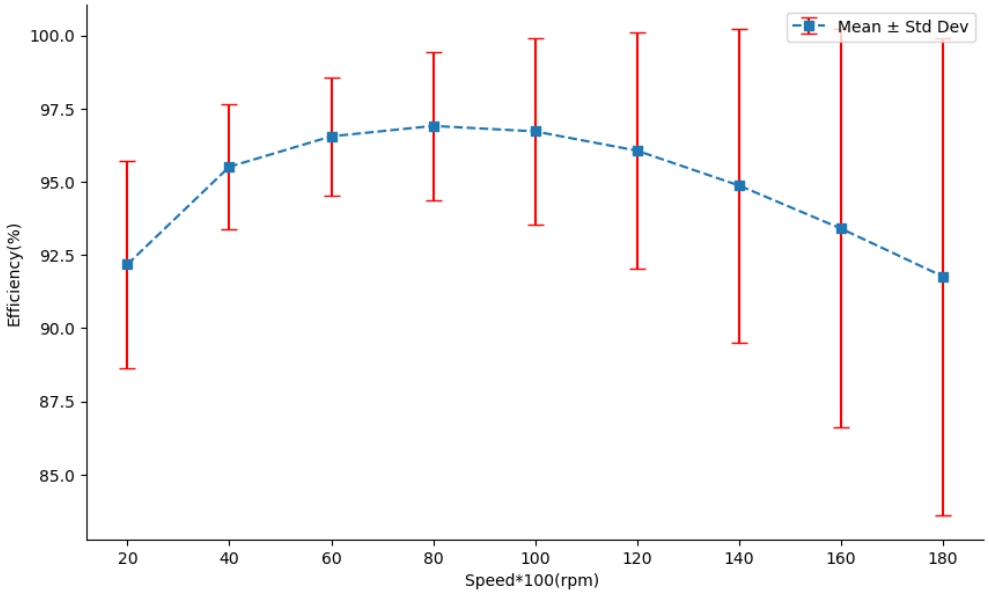



Figure 6.20: MLP Standard Deviation of ETA KPI Positive Grid across Speed Intervals

Figure 6.20 shows the standard deviation across different speed ranges for the MLP predictions. In comparisions to Figure 3.6 we infer that as the speed increases the target deviations of efficiency values are not so accurately captured by the model.

ETA KPI is harder to evaluate scoring as it is a 3D plot. Therefore, to visualize the Efficiency prediction deviation with its respective targets we do so for specific speeds across the entire torque range with Figure 6.21. The speeds are chosen at equal intervals of 2000 rpm.
We can observe that the most deviation occurs towards the higher torque values for each corresponding speed. In layman terms, this would be towards the ETA envelope tapering off.
This anomaly can be explained by the Mgrenz KPI predictions not being up to mark as it is responsible for trimming the edges of the envelope.
However it is not only the border of the envelope a question of concern here but also for neigbouring values of the envelope specifically from 1/4 the speed onwards. We donot have any way to teach the model this part of the grid as loss regularization because we cannot estimate the ennvelope dimensionality having already padded the targets to be of the maximum shape referred in 3.1.3. The reason why it deteroirates from 1/4 the speed onwards is because the ETA KPI's envelope starts to converge from around 6000 rpm as we can see in Figure 1.5.
Nevertheless, we see on average among the speeds, the RMSE is close to 5 which is 5% deviation from the target values as per Table 6.2.

Observations from the predictions helped to correct few discrepancies in our development for instance in the ETA grid we replaced 0 with NaN values which we later understood were both represented different in the grid.
As Efficiency values can take up values only between 0-100, we consider the same as constant across plots and use it as a baseline for determining the levels in the contour plot.
We have also left the output predictions for the Torque curve to remain as float values even when the target values are integers to preserve data precision. We give the client the flexibility to turn this on/off demand.

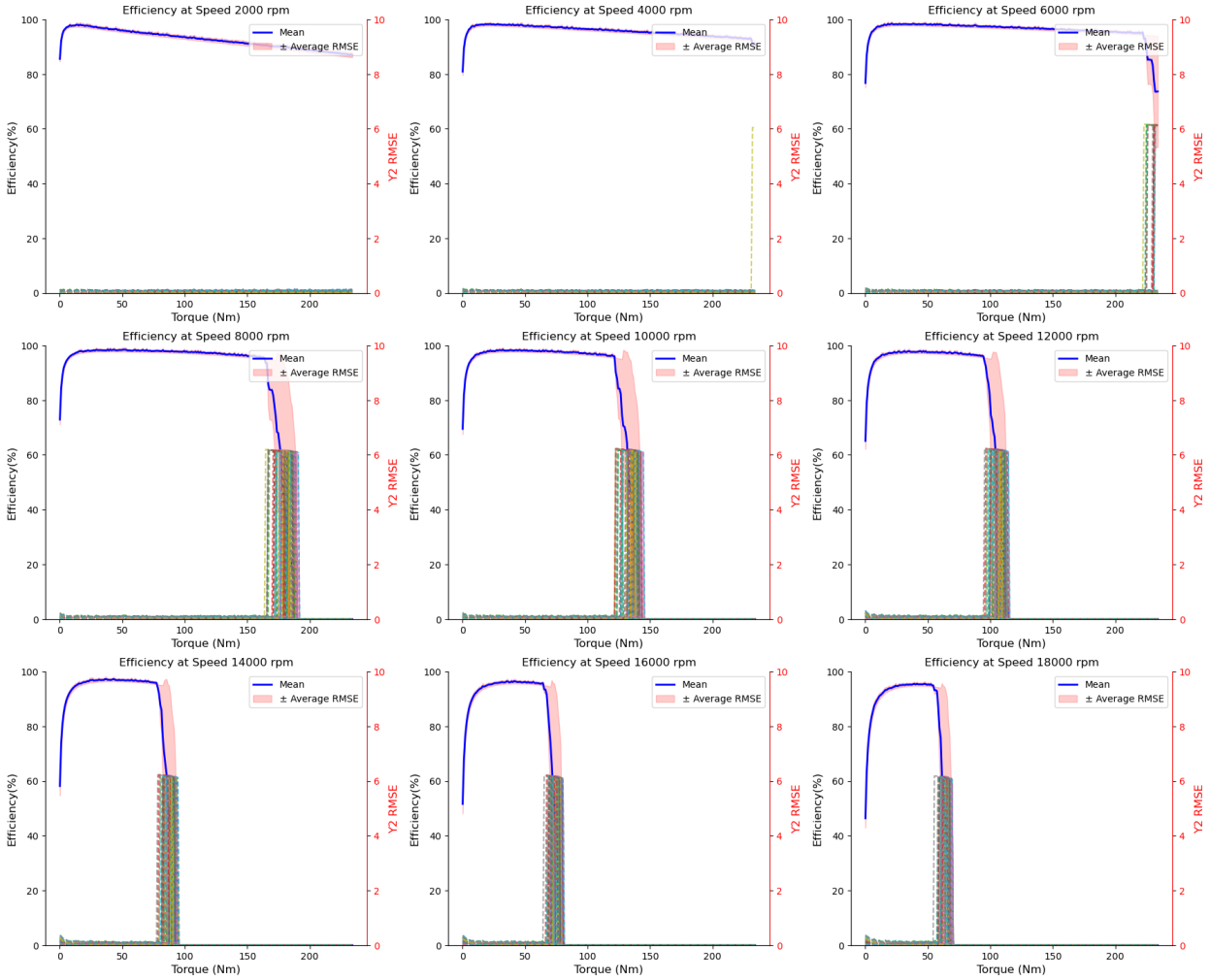



Figure 6.21: Eval MLP ETA RMSE KPI

Figure 6.22 shows the score statistics of the model performance of ETA KPI over the test dataset. The Y2 RMSE from Equation 5.12 is calculated for each sample and shown as a histogram.
On an average the RMSE is close to 13, which is 13% deviation from the target values as per Table 6.2.

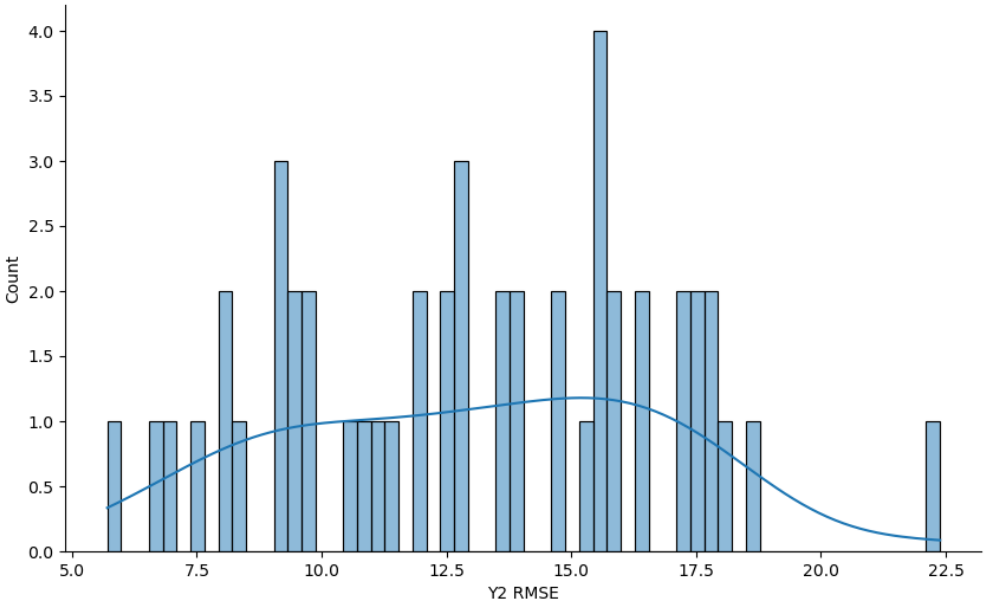



Figure 6.22: MLP Score statistics for ETA KPI

## 6.3 Results with Baseline

From our observations on how the predictions closely resembled that of the target values in Section 3.1, we have developed a Baseline model which is essentially the average of the train dataset.

### 6.3.1 Mgrenz KPI Results with Baseline

The Y1 Baseline score for the ETA KPI is formulated in Equation 6.2 :

$$\text{Y1 Baseline Score} = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\sqrt{\frac{1}{h}\sum_{j=1}^{h}(\bar{y}-y_{ij})^2}}_{Y1\ Baseline\ RMSE} \tag{6.2}$$

*where*

- $\bar{y}$ *: Baseline Average mean*
- $h$ *: Columns of 1D vector*
- *Y1 Baseline RMSE: RMSE for each test sample*

Figure 6.23 shows the Average RMSE and element wise RMSE for the test dataset performance with the Baseline Model.

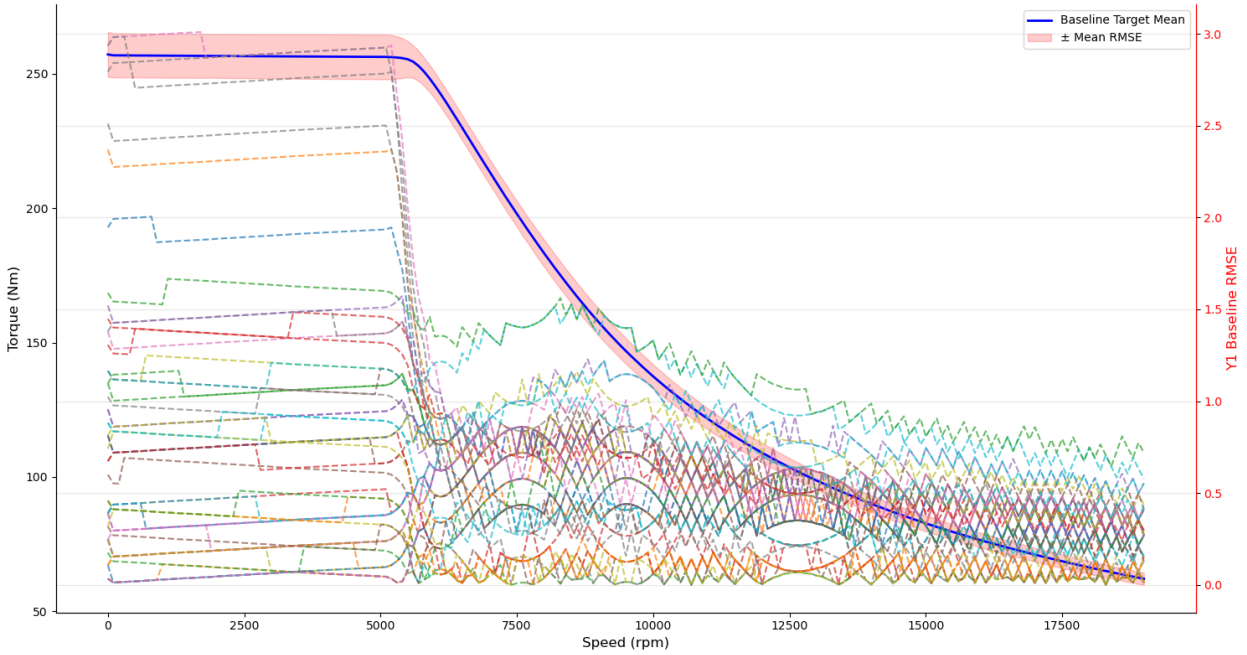



Figure 6.23: Baseline RMSE Evaluation for 2D KPI(Mgrenz)

We can see that the deviations are at its most towards the beginning of the curve. We also observe that there are no fluctuations in the curve as compared to the Predictions.

### 6.3.2 ETA KPI Results with Baseline

The Y2 Baseline score for the ETA KPI is formulated in Equation 6.3 :

$$\text{Y2 Baseline Score} = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\sqrt{\frac{1}{w}\frac{1}{h}\sum_{j=1}^{w}\sum_{k=1}^{h}(\bar{y} - y_{ijk})^2}}_{Y2\ Baseline\ RMSE} \tag{6.3}$$

*where*

- $w$ *: Rows of 2D vector*
- $h$ *: Columns of 2D vector*
- *Y2 Baseline RMSE : RMSE for each test sample*

Figure 6.24 gives a neat visualization of the Baseline Efficiency RMSE with its respective targets for specific speeds across the entire torque range. The speeds are chosen at equal intervals of 2000 rpm.
We infer from the plots that the targets have essentially the same values across the entire grid.
We arrive at this conclusion from the fact that the only region where we see significant deviation is at the ETA KPI envelope and we attribute this pattern to stem from the corresponding Mgrenz KPI's curve.
Thus, the efficiency values are almost the same for more than 3/4th of the grid across all samples.

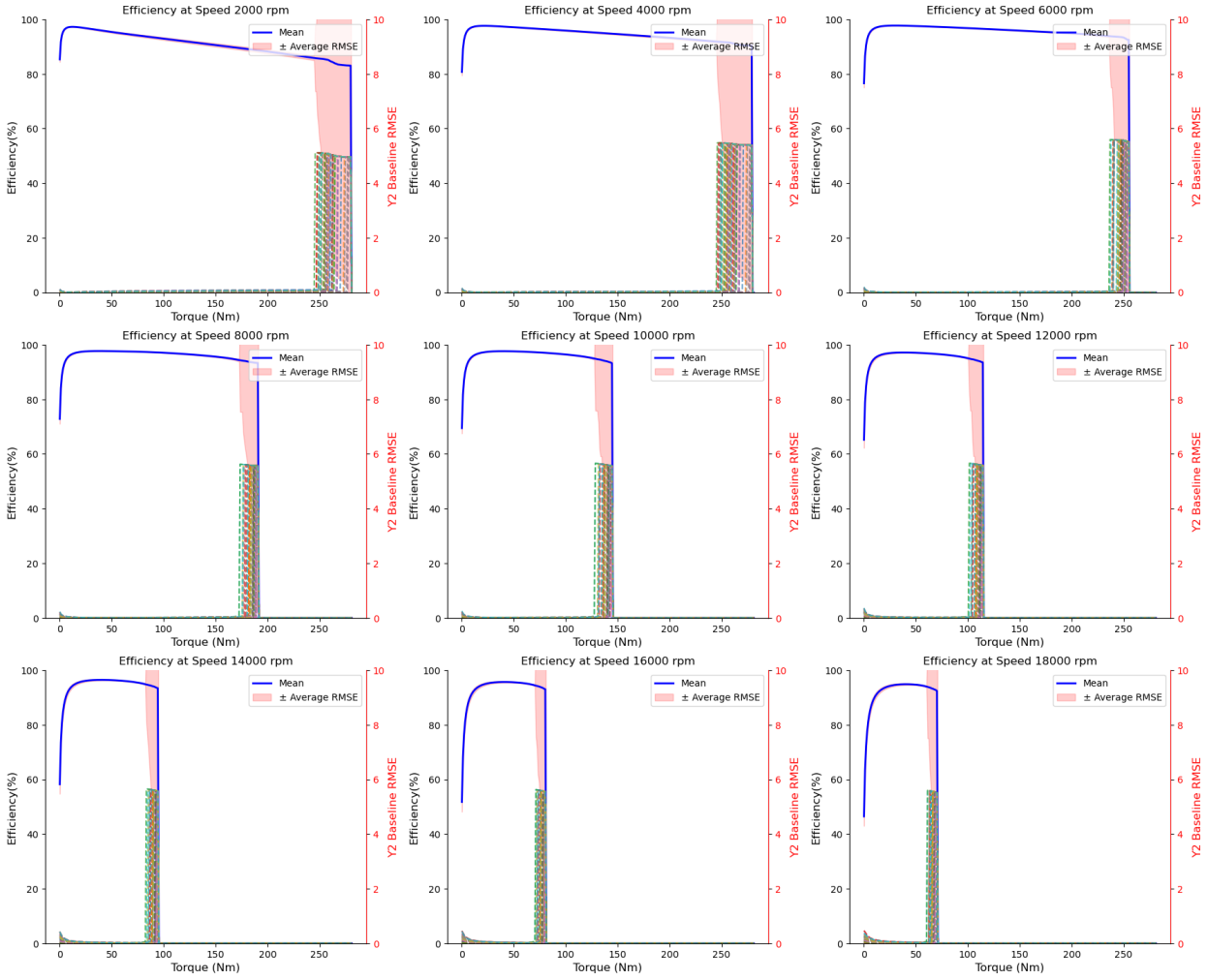



Figure 6.24: Eval Baseline ETA RMSE KPI

## 6.4 Results with Smoothening Loss Regularization

### 6.4.1 Mgrenz KPI Results with Smoothening Loss Regularization

Figure 6.25 shows the Average RMSE and element wise RMSE for the test dataset performance with the MLP Model with Smoothening Curve regularization discussed in Equation 5.2.
We can see that the deviations are at its peak towards the beginning of the curve. We also observe that there are fluctuations in the curve but relatively less when compared to Figure 6.14

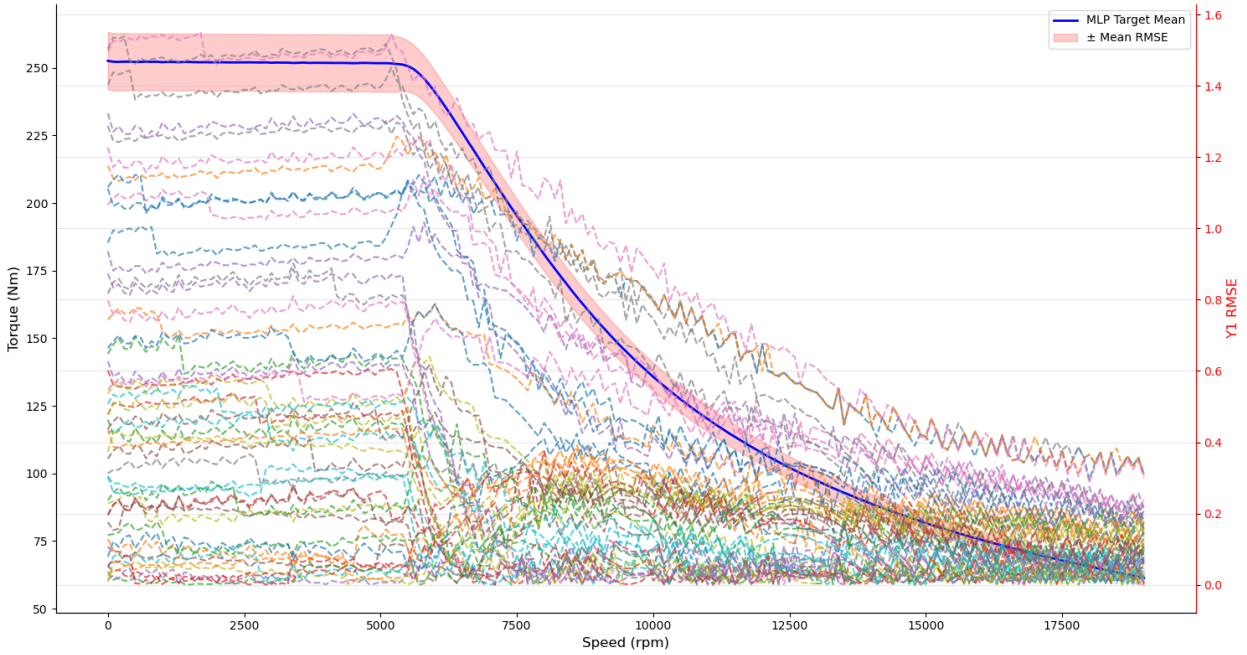



Figure 6.25: Smoothening Mgrenz RMSE Evaluation for 2D KPI(Mgrenz)

## 6.5 Results with No Loss Regularization

### 6.5.1 Mgrenz KPI Results with No Loss Regularization

Figure 6.26 shows the Average RMSE and element wise RMSE for the test dataset performance with the MLP Model without any regularizations.

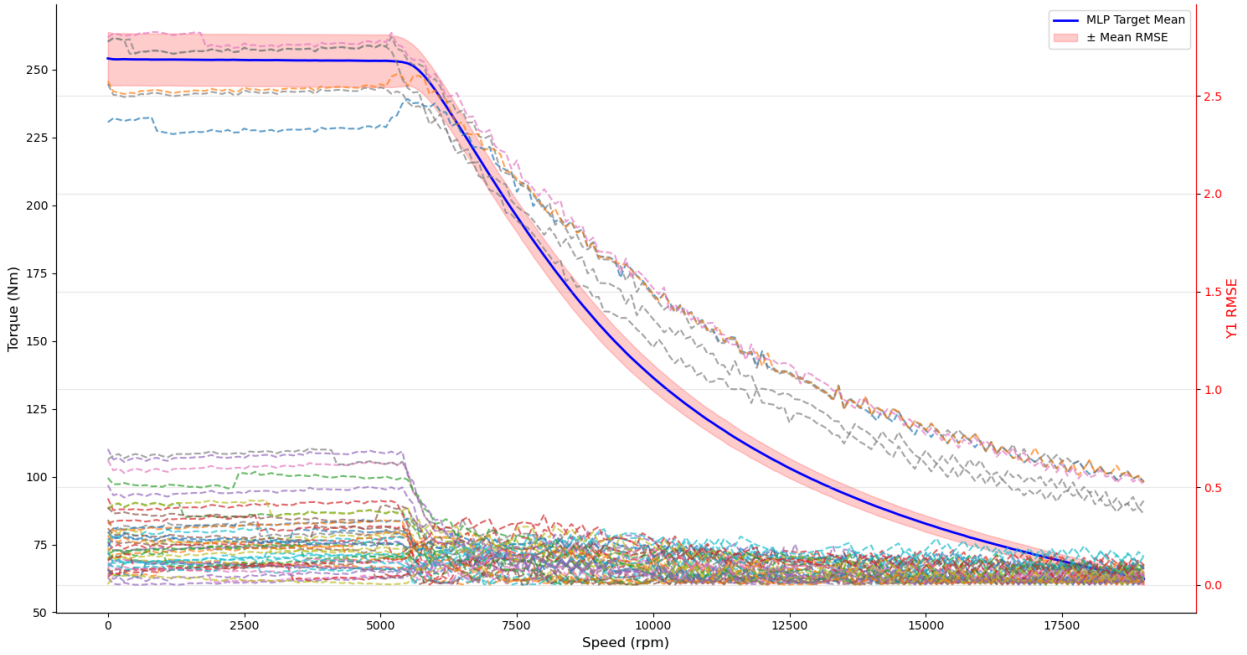



Figure 6.26: No Loss Regularization RMSE Evaluation for 2D KPI(Mgrenz)

We can see that the deviations are at its peak towards the beginning of the curve however the RMSE for most samples is close to negligible. We also note there are fewer fluctuations in the curve as compared to the Predictions with Loss Regularizations. This conclusion prompts us to omit the loss regularization for the Mgrenz KPI's.

### 6.5.2 ETA KPI Results with No Loss Regularization

Figure 6.27 gives a neat visualization of the Baseline Efficiency RMSE with its respective targets for specific speeds across the entire torque range. The speeds are chosen at equal intervals of 2000 rpm.
We observe similar patterns to the ones generated with loss regularization in place. This fuels us to also omit the loss regularizations for the ETA KPI's.

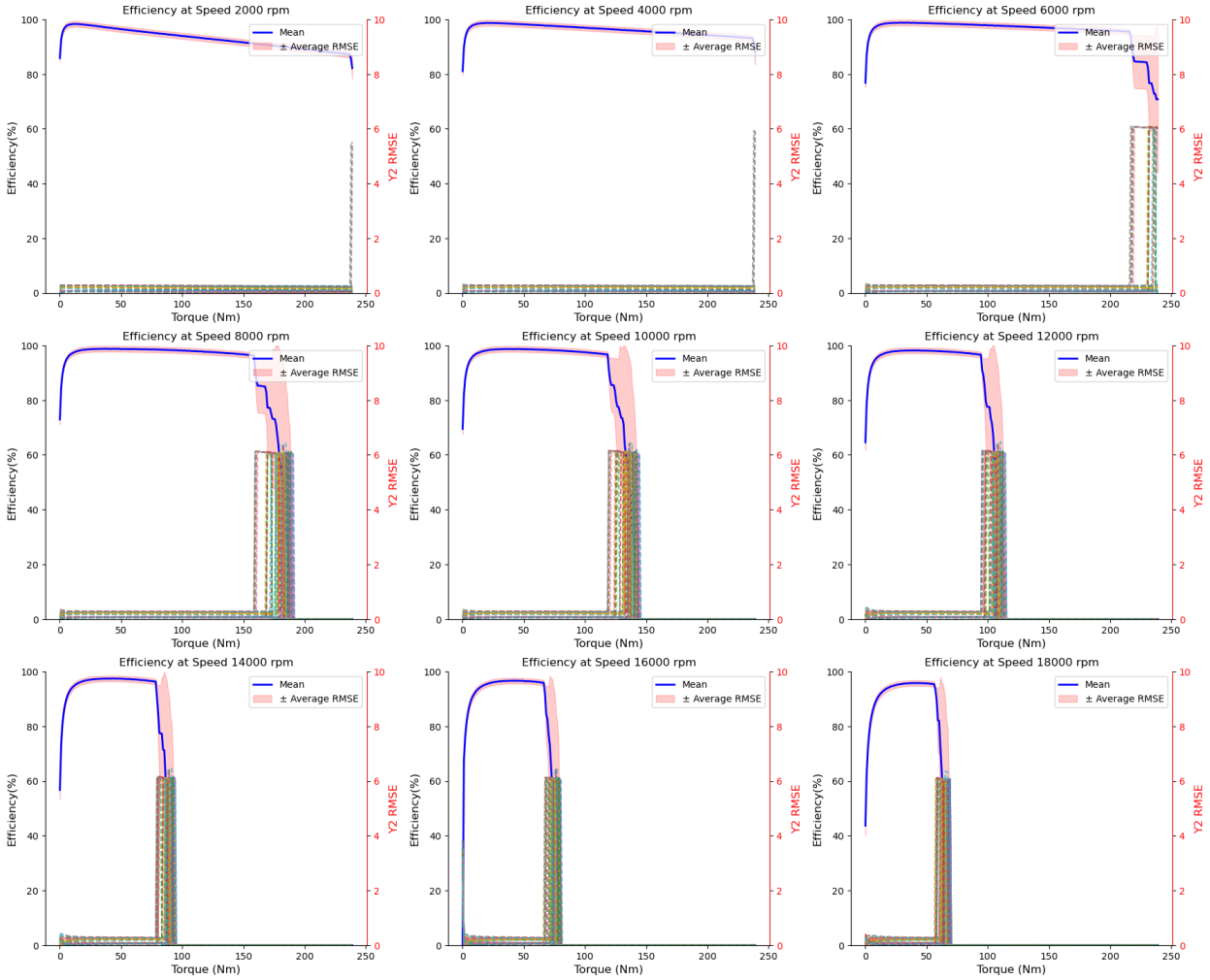



Figure 6.27: Eval No Loss Regularization ETA RMSE KPI

## 6.6 Ablation Studies

We conduct comprehensive ablation studies on the MLP model with different loss regularizations and the Baseline model for both our targets.

| Model | Y1 Score | Y2 Score |
|---|---|---|
| Baseline | 4.8201 | 15.3708 |
| MLP[a] | 6.0652 | 13.0469 |
| MLP[b] | 5.5052 | 15.0257 |
| MLP[c] | 4.9887 | 12.6612 |

Table 6.3: Ablation Studies

[a]MLP With Decreasing Y1 Loss Regularization
[b]MLP With Smoothening Y1 Loss Regularization
[c]MLP without both Y1 and Y2 Loss Regularization

Results in Table 6.3 indicate that our MLP model without loss regularization's performance is in par with the Baseline model on the Mgrenz KPI and has outperformed it on the ETA KPI.

This could be because the regularizations are restrictive and discourage the model to learn better.
the enriched ablation studies we undertook demonstrate the robustness of our method across varying hyper-parameter setings.
The inferences shown are strictly speaking from the Double V Magnet Topology which assumed the bulk of the data we had received.

Considering the limitations in dataset, conducting thorough examinations of a substantial volume is typically unfeasible. Therefore, the primary purpose of the thesis is to study the possibility of modelling the KPI's predictions and our learnings from it.

We have used Python 3.10.14 for our development and the Pytorch library compatible with Cuda.
The model was trained on a NVIDIA Tesla V100 GPU with 32 GB Memory.
Source code is available at the GitHub Repository.

# Chapter 7

# Conclusion

This thesis offers a fresh outlook to the possibility of modelling the performance of an EM using GNNs. We have developed a topology invariant MLP model to predict 2 KPI's of an EM from its parameteric requirements. It would be very beneficial to EM manufacturers to understand the operating range of the vehicle using the motor and make calculated assumptions of whether to manufacture it.
This work enables EM designers to generate KPI's at almost both negligible costs and minimal time. Thus offering an escape from the cons of using FEM simulations.
It also lays the foundation for future work on being able to generate electric motor design parameters conditioned on the 2 KPIs we predicted. We envision that our study will provide new perspectives to researchers engaging in the field of electric motor design and hope our contribution will aid in the same.

## 7.1 Future Improvements

I would suggest the following improvements to our study :

1. Build a model that uses the Mgrenz KPI prediction to predict its corresponding ETA KPI.
2. Although we have designed the model to be topology invariant for 3 topologies, we only had sufficient data from Double V Topology to draw evaluations from. Further evaluations on the other topologies would be beneficial to critically assess our model's performance.
3. Another interesting study would be to benchmark model performance when one backpropagates the losses for each KPI individually rather than weighing them.
4. Additionally, the motivation of building a topology invariant model was the reason we have considered building a heterogeneous graph to model the data. The machinery is elaborated in Section 4.4. Such a model could serve as yet another ablation study to our problem.

# List of Figures

# List of Tables

# Bibliography


[01] Tsai Hor Chana, Guosheng Yina, Kyongtae Baeb, Lequan Yu *Multi-task Heterogeneous Graph Learning on Electronic Health Records.* cs.LG, 14 Aug 2024.

[02] Xiaocheng Yang, Mingyu Yan, Shirui Pan, Xiaochun Ye, Dongrui Fan *Simple and Efficient Heterogeneous Graph Neural Network.* cs.LG, 1 Sep 2023.

[03] Xinru Huang *A Heterogeneous Graph Neural Network Model for Electric Vehicle Purchase Propensity Prediction.* ICDSCA, Oct 2023.

[04] Qingsong Lv, Ming Ding, Qiang Liu, Yuxiang Chen, Wenzheng Feng, Siming He, Chang Zhou, Jianguo Jiang, Yuxiao Dong, Jie Tang *Are we really making much progress? Revisiting, benchmarking, and refining heterogeneous graph neural networks.* cs.LG, 30 Dec 2021.

[05] Fredrik Johannessen; Martin Jullum *Finding Money Launderers Using Heterogeneous Graph Neural Networks.* ICDSCA, 25 July 2023.

[06] Damian Owerko, Fernando Gama, Alejandro Ribeiro *Predicting Power Outages Using Graph Neural Networks.* cs.LG, Nov 2018.

[07] Mikko Tahkola, Janne Keränen, Denis Sedov, Mehrnaz Farzam Far, Juha Kortelainen *Surrogate Modeling of Electrical Machine Torque Using Artificial Neural Networks.* ICDSCA, Dec 17 2020.

[08] AKM Khaled Ahsan Talukder, Bingnan Wang and Yusuke Sakamoto *Electric Machine Two-dimensional Flux Map Prediction with Ensemble Learning.* ICEMS, 2022.

[09] Kazuhisa Iwata, Hidenori Sasaki, Hajime Igarashi, Daisuke Nakagawa, and Tomoya Ueda *Generalization Performance in Predicting Torque Characteristics Using Convolutional Neural Network and Stator Magnetic Flux.* ICDSCA, 3 Mar 2024.

[10] Simone Ferrari, Paolo Ragazzo, Gaetano Dilevrano, Gianmario Pellegrino *Flux-Map Based FEA Evaluation of Synchronous Machine Efficiency Maps.* WEMDCD, 2021.

[11] Vivek Parekh, Dominik Flore, Sebastian Schöps *Deep Learning-Based Prediction of Key Performance Indicators for Electrical Machines.* Jan 22, 2021.

[12] Carlos Candelo-Zuluaga, Antonio Garcia Espinosa, Jordi-Roger Riba, Pere Tubert Blanch *Computationally Efficient Analysis of Spatial and Temporal Harmonics Content of the Magnetic Flux Distribution in a PMSM for Efficiency Maps Computation.* 2020.

[13] Yihao Xu, Bingnan Wang, Yusuke Sakamoto, Tatsuya Yamamoto, and Yuki Nishimura *Comparison of Learning-based Surrogate Models for Electric Motors.* COMPUMAG, 2023.

[14] Marius Stender, Oliver Wallscheid, Joachim Böcker *Accurate Torque Estimation for Induction Motors by Utilizing a Hybrid Machine Learning Approach.* PEMC, 2021.

[15] Katsuyuki Narita, Hiroyuki Sano, Nicolas Schneider, Kazuki Semba, Koji Tani, Takashi Yamada, Ryosuke Akaki *An Accuracy Study of Finite Element Analysis- based Efficiency Map for Traction Interior Permanent Magnet Machines*: 2020.

[16] Jo Asanuma, Shuhei Doi, Hajime Igarashi *Transfer Learning Through Deep Learning: Application to Topology Optimization of Electric Motor.* 3 Mar 2020.

[17] Diederik P. Kingma, Jimmy Ba *Adam: A Method for Stochastic Optimization.* 30 Jan 2017.

[18] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, Philip S. Yu *A Comprehensive Survey on Graph Neural Networks.* 04 Dec 2019.

[19] Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Caglar Gulcehre, Francis Song, Andrew Ballard, Justin Gilmer, George Dahl, Ashish Vaswani, Kelsey Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matt Botvinick, Oriol Vinyals, Yujia Li, Razvan Pascanu *Relational inductive biases, deep learning, and graph networks.* 17 Oct 2018.

[20] Shichao Zhu, Chuan Zhou, Shirui Pan, Xingquan Zhu, Bin Wang *Relation Structure-Aware Heterogeneous Graph Neural Network.* 2019.

[21] Xingtong Yu, Yuan Fang, Zemin Liu, Xinming Zhang *HGPROMPT: Bridging Homogeneous and Heterogeneous Graphs for Few-shot Prompt Learning.* 5 Jan 2024.

[22] Qimai Li, Zhichao Han, Xiao-Ming Wu *Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning.* 22 Jan 2018.

[23] Ziniu Hu, Yuxiao Dong, Kuansan Wang, Yizhou Sun *Heterogeneous Graph Transformer.* 3 Mar 2022.

[24] Yiling Zou, Jian Shu *Heterogeneous Network Node Classification Based on Graph Neural Networks.* IEEE, 2023.

[25] Joshua Melton, Siddharth Krishnan *muxGNN: Multiplex Graph Neural Network for Heterogeneous Graphs.* IEEE, 9 Sept 2023.

[26] Carl Yang, Yuxin Xiao, Yu Zhang, Yizhou Sun, Jiawei Han *Heterogeneous Network Representation Learning: A Unified Framework with Survey and Benchmark.* cs SI, 17 Dec 2020.

[27] Justin Gilmer 1 Samuel S. Schoenholz 1 Patrick F. Riley 2 Oriol Vinyals 3 George E. Dahl *Neural Message Passing for Quantum Chemistry.* cs SI, 12 June 2017.

# Declaration on oath

I hereby certify that I have written my master thesis independently and have not yet submitted it for examination purposes elsewhere. All sources and aids used are listed, literal and meaningful quotations have been marked as such.

---

Lilly Abraham K64889, 11.12.2024

# Consent to Plagiarism Check

I hereby agree that my submitted work may be sent to PlagScan (www.plagscan.com) in digital form for the purpose of checking for plagiarism and that it may be temporarily (max. 5 years) stored in the database maintained by PlagScan as well as personal data which are part of this work may be stored there.

Consent is voluntary. Without this consent, the plagiarism check cannot be prevented by removing all personal data and protecting the copyright requirements. Consent to the storage and use of personal data may be revoked at any time by notifying the faculty.

---

Lilly Abraham K64889, 11.12.2024