National College of Ireland

A Graph Neural Network for Predicting the Magnetic Interaction Between Atoms

MSc Research Project
MSc in Data Analytics

Jaswinder Singh
Student ID: x19219997

School of Computing
National College of Ireland

Supervisor: Dr. Paul Stynes
Dr. Pramod Pathak

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Jaswinder Singh
Student ID:	x19219997
Programme:	MSc in Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Dr. Paul Stynes , Dr. Pramod Pathak
Submission Due Date:	16/08/2021
Project Title:	A Graph Neural Network for Predicting the Magnetic Interaction Between Atoms
Word Count:	6585
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	20th September 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

A Graph Neural Network for Predicting the Magnetic Interaction Between Atoms

Jaswinder Singh
x19219997

Abstract

The magnetic interaction between a pair of atoms can be determined by calculating the value of the quantity known as the scalar coupling constant (SCC). The SCC plays a crucial role in the analysis of 3D structure of organic matter and its precise calculation can be used in a variety of tasks like drug discovery, toxicity determination, etc. The quantum mechanical density functional theory (DFT) provides a theoretical framework for predicting the magnetic interactions (or SCC) between the atoms. The quantum mechanical computations use the 3D structural information of the molecule as an input for precise calculation of SCC. However, these computations are extremely time consuming and computationally expensive. To compute SCC efficiently and accurately, a novel graph neural network (GNN) is proposed, combining the message passing elements from the Message Passing Neural Networks (MPNN) model with the multi-head attention layers as in transformer encoder. The proposed model is named as the *Message Passing Molecular Transformer (MPMT)*. Different bond level and atom level features like cosine angles and dihedral angles were created using the RDkit package, to improve the predictions of the model. To demonstrate and validate the superiority of the proposed model, the CHAMPS dataset consisting of structural information and coupling constant values of around 10 million different molecules collected by University of Bristol was used. The results were evaluated on the basis of the log MAE(mean absolute error) values for each coupling type and the final score for the model was computed by averaging over all the coupling types. The MPMT model was able to achieve a final score of -2.873 which corresponds to the mean absolute error of 0.0565 Hz whereas the MPNN was able to achieve an average final score of -2.19 corresponding to the MAE of 0.111 Hz. Our proposed model was able to outperform the state-of-the-art results for the CHAMPS dataset (Jian et al. (2020)) which is MAE value of 0.096 Hz. This study will eventually benefit to various domains; for e.g it will contribute towards efficient and drug development for different diseases, it can be utilised to improve the current NMR techniques for molecular property prediction and can also be used in materials science industry for enhancing cosmetics' quality. It will also greatly benefit researchers in studying the molecular structure more efficiently, saving lot of research costs.

1 Introduction

Two atoms interact magnetically when the magnetic fields produced by their subatomic units (protons and electrons) are influenced by each other. A scalar coupling constant (SCC), usually denoted by J is a measure of the interaction between two protons. The molecular interactions are a great and powerful resource for molecular discoveries. The networks describing molecular interactions can be found today in almost every biological systems. They have enhanced our understanding of different biological systems and have enabled us to explore previously unmapped interactions between certain entities like drug-target interactions, protein-protein interactions, gene-disease interactions, etc. The precise knowledge of molecular interactions can also be used for determining the structure of a chemical compound which in turn can be used in determining its various properties like solubility, conductivity, electrostatic potential, binding affinity, etc. Structure of a compound also greatly influences it behaviour under different environments. The Nuclear Magnetic Resonance (NMR) is a popular technique widely used for obtaining the physical and chemical information about the structure of molecules. *Scalar coupling constant (SCC)* is one of the two important aspects in NMR spectroscopy besides *chemical shift*.

The current methods for predicting the magnetic interactions (or SCC) between atoms use quantum mechanical computations which use only the 3 dimensional structural information of the molecule as an input. A quantum mechanical framework called the Density Functional Theory (DFT) is used for carrying out these computations (Gryff-Keller and Szczeciński (2016), Komorovsky et al. (2020)). But these methods are extremely computationally expensive and require weeks for performing predictions

for a single molecule. The search for a robust, faster and a reliable algorithm/model for predicting these interactions is still an active area for research. The graph networks have proven to be a promising candidate for such models. Over the last few years, researchers have come up with various graph network based frameworks and accomplished state-of-the-art results for different molecular property prediction tasks. But there is only a handful of literature involving prediction of coupling constants between the atoms. This research proposes a novel graph based network which is a combination of two different architectures - a transformer encoder and a message passing neural network.

The aim of this research is to investigate to *what extent a novel geometric deep learning approach based on graph neural networks (GNN) can be used to predict the magnetic interaction between the atoms in a molecule ?* To address the research question, the following sets of specific research objectives were derived.

- To design:
  - A Message Passing Neural Network (MPNN) based on the work of Gilmer et al. (2017) to predict the coupling constant between atoms.
  - A custom graph neural network *Message Passing Molecular Transformer* consisting of elements from the MPNN (Gilmer et al. (2017)) and the transformer encoder (Vaswani et al. (2017)) for the precise prediction of magnetic interaction (values of SCC) between atomic pairs.
- To demonstrate the implementation of the above two architectures on the CHAMPS dataset available on Kaggle.
- To evaluate the performance of the above models on the basis of log mean absolute error (MAE) and root mean squared error (RMSE) averaged over all coupling types present in the CHAMPS dataset.

The major contribution of this research is a novel graph based network called Message Passing Molecular Transformer (MPMT) for predicting the magnetic between the different atomic pairs. It uses the multi-head attention layers as in transformer encoder (Vaswani et al. (2017)) as the building blocks and employs the message passing elements from the Message Passing Neural Network (MPNN) model (Gilmer et al. (2017)). The training time and results for the proposed model will be compared against the other state-of-the-art machine learning methods that have been previously applied on similar datasets for molecular property prediction tasks. This study will aid the researchers in performing computationally expensive quantum computations for molecules efficiently and faster. This research will subsequently contribute towards development of new drugs to fight diseases more effectively, study of various protein based biological systems, carrying out different cellular tasks, developing new cosmetics and will save lots of research costs.

The remainder of this paper is organised as follows:

- Section 2: This section contains a comprehensive overview of the existing literature in the domain of molecular property prediction and graph neural networks. The section has been further divided into several relevant subsections based on the specified criterion.
- Section 3: This section explains the overall methodology followed to obtain the results in detail. It contains several subsections corresponding to each step in the overall methodology, that explains and justifies the procedure followed for that particular step.
- Section 4: This section lays out the underlying architecture for the implementation of the proposed novel graph network. It also contains the description of the proposed Message Passing Molecular Transformer (MPMT) model.
- Section 5: This section discusses the implementation details of the proposed model. It highlights various tools and softwares used in the preparation of the model, and the final stages of the model development i.e the outputs produced, transformed data and the questionnaires administered.
- Section 6: This section discusses the details of different experiments performed and their results. It is also divided into separate subsections for each experiment. It also contains a discussion section which summarizes the results of each experiment.
- Section 7: This section discusses the concluding remarks and the prospects of future work in the domain.

2 Related Works

Before moving onto the methodology for our research, it is important to critically review and analyze the existing literature in the domain. This helps us analyze the limitations and shortcomings of the previous works and gain insights into how they can be improved upon.

The prediction of molecular properties (like chemical shift and SCC) is an essential step in tasks like development of new drugs, studying the structure of unknown compounds, etc. Strength of magnetic interaction between the atoms is one such property measured by determining the corresponding scalar coupling constant (SCC) for the atomic pair. There are different types of SCC based on the types of atoms involved. There have been several studies over the past few years using a variety of machine learning algorithms like XGBoost, LightGBM, etc. for molecular property prediction tasks (like drug-target affinity prediction, toxicity determination, etc.). This section summarizes the key studies in the domain of molecular property prediction (including prediction of scalar coupling constant). It has been divided into the following subsections:

- *Brief Review of Graph Based Networks*: This subsection gives a brief overview of the development of various graph based networks like Graph Convolutional Network (GCN), Graph Attention Network (GAT), etc.
- *Molecular Property Prediction*: This section contains the critical review of the key literature that is the backbone of this study along with the other essential studies in the domain of molecular property prediction. It has been further divided into following subsections:
  - *Using Machine Learning Frameworks*: Summarizes the literature which employ machine learning algorithms for molecular property prediction tasks.
  - *Using Graph Based Networks*: Summarizes the literature which use graph based networks for predicting the molecular properties.

2.1 Brief Review of Graph Based Networks

In layman's terms, a graph is a network that helps define and visualize the relationships between various components. In more formal terms, the components are known as the vertex/nodes and the relationships between them are known as the edges. Graph theory is all about the study of the properties of these types of networks and how they can be used to model and solve a whole host of interesting problems. Recently, the studies of using machine learning algorithms to analyze graphs have been getting more attention because of the ability of graphs to model and express complex real life structures like social networks (Hamilton et al. (2017)), physical systems (Battaglia et al. (2016)), protein based biological networks (Fout (2017)) and many more. The increase in the use of graph based networks in a wide range of applications is because of the tremendous expressive power of graphs resulting in high performance for a variety of machine learning tasks.

The first use of graph neural networks (GNN) can be traced back to the year 2005. The major motivation for GNNs lies in the convolutional Neural Networks (CNNs). The local connections and the shared weights are the most important keys behind the CNNs ability to extract localized features and compose them together to construct extremely expressive representations. These properties are inherited by the graphs quite well and has been the primary reason for their computational efficiency as compared to the other machine learning algorithms. Zhang, Cui and Zhu (2020) provide the most comprehensive and up-to-date review of the graph networks. They divide graph neural networks into 5 groups based on their architecture and training strategy: Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), Graph auto-encoders, graph generative networks and graph spatial temporal networks. They also discuss the various potential applications of graph networks and potential future research work that could be performed.

Hechtlinger et al. (2017) introduced the graph convolutional networks (GCNs) for the first time in the year 2017. GCN was introduced as a generalization of the CNNs having the capability to operate on graph structured data in addition to the low-dimensional grid data such as images. They demonstrated their proposed architecture on the MNIST and the Merck Molecular activity dataset. They extend their discussion by concluding that the convolution method can be applied even to the usual classification and regression problems by learning the underlying graph structure through correlation matrix or other methods.

The major limitation of the GCNs is that they are only designed to learn on the fixed and homogeneous types of graphs. This limitation was overcome by the introduction of Graph Transformer Network

(GTNs) in the year 2019 by Yun et al. (2019). The GTNs have the ability to produce new graph structures and identify any useful connections between the unconnected nodes in the original graph. Their study shows that GTNs are able to learn on new graph structure based on the given data and task without any existing knowldege about the domain. They demonstrated the performance of their proposed model on three benchmarked node classification datasets and were able to achieve the state-of-the-art results on all of them.

2.2 Molecular Property Prediction

In the previous section, we have seen that the graph neural networks emerged as a way to efficiently deal with the graph structured data but they have successfully paved their way across many domains over the years. The reason for the success of the graph based networks across diverse domains is the ability to learn on complex structured data and their computational efficiency due to their sharing weights property. We have also seen that some of these networks perform extremely well on molecular graphs (Coley et al. (2017), Kearnes et al. (2016)) resulting in their wide usage in the domain of molecular property prediction. This section critically review some of the key literature in the application of graph based networks in molecular property prediction. It also highlights the works which inspired our current research. Section 2.2.1 summarizes various machine learning frameworks that have been used for various molecular property prediction tasks. Section 2.2.2 is at the heart of this study and discusses the key literature in which graph networks have been used in the prediction of molecular property. It also highlights the studies which inspired the underlying architecture of proposed MPMT (Message Passing Molecular Transformer) model.

2.2.1 Using Machine Learning Frameworks

The most recent work on the CHAMPS dataset(which is used by our research) for predicting the coupling constant was done by Zhang, Deng and Jia (2020). They used two tree based gradient boosting algorithms namely XGBoost and LightGBM for predicting the SCC (scalar coupling constant). They used the 3D coordinates of the atoms in the molecule to find the objective function able to map relationship between the features and the target variable i.e SCC. Their best result on the 90:10 train-test split data has the MAE of 4.34, an R^2 value of 0.93 and RMSE value of 8.65. They were even able to achieve an R^2 value of 1 for some of the molecules. From plotting the learning curve they were able to conclude that although, the error for training and test set continuously reduced and attained a specific value at the end, this value was less for LightGBM which leads to the conclusion that LightGBM is the better performing algorithm. The major limitation of their work is that the predictions they obtained deviate from the DFT calculations.

Another research on the same dataset was done by Shibata and Kaneko (2021). They also used the LightGBM algorithm for the task. Their study heavily depends on the feature engineering process. They propose several descriptors like hybridization, aromaticity, Euclidean Distance, etc to build regression models of the form $y = f(X)$ between the descriptors X and the NMR results y, from which the y-values can be determined from the X values of samples whose y-values are unknown. They obtained the best R^2 and RMSE values of 0.986 and 1.25 respectively for the 1JCH coupling type. They were able to outperform the traditional method in which the descriptors are RDkit descriptors for all the coupling types. But their algorithm also suffered from the same limitation as Zhang, Cui and Zhu (2020), i.e the less predictive accuracy as compared to DFT calculations.

The lightGBM algorithm has also been able to perform well in other molecular property prediction tasks like drug affinity prediction. In an interesting research done by Pu et al. (2019), 4 different algorithms namely LightGBM, XGBoost, CNN and Multi Layer Perceptron (MLP) were compared in a drug affinity prediction task on the dataset containing around 160k samples. They used the RMSE and R^2 as the evaluation metrices. They found out that the gradient boosting algorithms namely XGBoost and LightGBM were able to outperform the CNN and MLP

2.2.2 Using Graph Based Networks

Graph based networks have emerged as a computational efficient method when it comes to predicting various molecular properties. The conventional machine learning algorithms, boosting algorithms like LightGBM in particular, have shown promising results in predicting the molecular properties like the scalar coupling constant which means they can be also be used for other related tasks as well. Although being significantly faster, the results produced by the algorithms deviate significantly from the quantum

mechanical methods like DFT, which are considered state-of-the-art. This section highlights some of the important literature where graph based networks have been used for molecular property prediction tasks ranging from coupling constant prediction to drug binding affinity prediction to protein-protein interaction determination.

The most recent work using graphs for predicting the magnetic interaction was done by Jian et al. (2020). The authors propose a novel architecture called graph embedding local self attention encoder (GELAE). They use different features like bond angle, bond length and dihederal angle to represent the coupling system in a structure invariant way and then use the self attention module along with the adjacency matrix to effectively extract various features of the system. They used the classification based loss function and demonstrated their architecture on CHAMPS dataset. They were able to achieve an MAE value of 0.0963 Hz.

Xiong et al. (2019) proposed a novel GNN architecture called *Attentive FP* which uses the graph attention mechanism to learn from drug discovery dataset. Their architecture consists of a fully connected input layer, two different graph attention modules - one containing 3 layers for atom embedding and the other containing 2 layers for molecule embedding followed by a fully connected task later. They test their model on the MoleculeNet and the QM9 datasets and were able to achieve state-of-the-art results for various tasks. Their model achieved the lowest RMSE on all benchmarked physical chemistry datasets.

The major motivation for the current study comes from the work of Gilmer et al. (2017) proposing a novel framework called Message Passing Neural Networks (MPNN). Through MPNN, they proposed an efficient method through which the model can learn the message passing algorithm and an aggregation method for computing the function of the whole input graph. They demonstrated the efficiency of their model on various different molecular property benchmarked datasets including QM9. The authors were able to show that MPNNs with appropriate update, message and output functions can easily outperform baseline models eithout having the need for performing any feature engineering. They also described the different variants of the MPNN model, each with a slightly different method of learning and input function. They also throw some light on designing the MPNNs for larger graph networks.

The other half of the MPMT architecture proposed in this research utilises the multi-head attention layers, same as described by Vaswani et al. (2017) in their ground breaking research. The idea of building a blenedd model from MPNN and transformer was inspired from the study conducted by Withnall et al. (2020) in which they introduced attention and edge memory elements to the existing MPNN architecture. They proposed three different MPNN architectures variations - SELU-MPNN, using the SELU activation fucntion with the standard MPNN. Attention Message Passing Neural Network (AMPNN), using the attention mechanism for computing the aggregate function and EMNN (Edge Memory Neural Network), which uses the aggregate information from the neighbouring atoms to compute the message for central atom. They demonstrate the performance of their models as part of data pre-processing, in addition to using the benchmarked MolNet and QM8 datasets for various standard classification tasks.

Jørgensen et al. (2018) following the works of Gilmer et al. (2017) proposed another variation of MPNN with an edge update network which allows the information exchange between the atoms depend on the state of the receiving atom. They tested their model on three benchmarked datasets namely QM9, the materials project and OQMD. They trained their model with Adam optimizer and an initial learning rate of 5×10^{-4} and were able to achieve the lowest MAE for all the three datasets (10.5 for QM9, 22.7 for Mat. Proj. and 14.9 for OQMD).

2.3 Conclusion

After a critical review of the aforementioned studies, we can conclude that among the existing techniques for SCC prediction, the graph based neural networks like Graph Neural Networks (GNNs), Graph Attention Networks (GATs) and different variations of the Message Passing neural Networks (MPNNs) were found to outperform the traditional machine learning algorithms like LightGBM and XGBoost, by a significant amount. This can be affirmed by comparing the log MAE values of different models. Particularly, for the CHAMPS dataset, the state-of-the-art MAE value of 0.0963 was obtained by Jian et al. (2020) using a novel GELAE (Graph Embedding Local Self-Attention Encoder) framework. Other machine learning approaches like XGBoost, Light GBM were adopted by Zhang, Deng and Jia (2020) who were able to achieve an overall MAE of 4.34, averaged over all the coupling types. Another study using LightGBM was conducted by Shibata and Kaneko (2021) in which various new features were engineered to determine the SCC. They were able to RMSE value of 1.25, outperforming the previous study done by Zhang, Deng and Jia (2020). Although, all of the above models perform well for predicting SCC, the power and efficiency of graph networks remain unexplored. The major limitation of all of

the machine learning approaches is the significant difference from the values computed by the quantum mechanical/experimental methods. The lack of useful features in the models to produce significantly accurate results is another shortcoming of the traditional machine learning approaches. The state-of-the-art GELAE framework is a graph based architecture and it explores to some extent, how we can utilise the expressive power of graphs is such tasks, while also giving some useful directions for future work. Therefore, in this study, we explore a graph neural network combining the elements of a MPNN and a transformer for predicting the SCC values between atoms. This choice of architecture has been inspired after carefully analyzing the existing methods and their shortcomings and advantages.

3 Methodology

The methodology is one of the most essential part of a research. It is important to follow the correct methodology for the research problem chosen as it essential for the successful implementation of the proposed ideas leading to successful completion of the research objectives. For this research, the methodology can be divided into the steps described below (refer to figure 1). The following steps can be followed in order to replicate the results obtained by our proposed model.

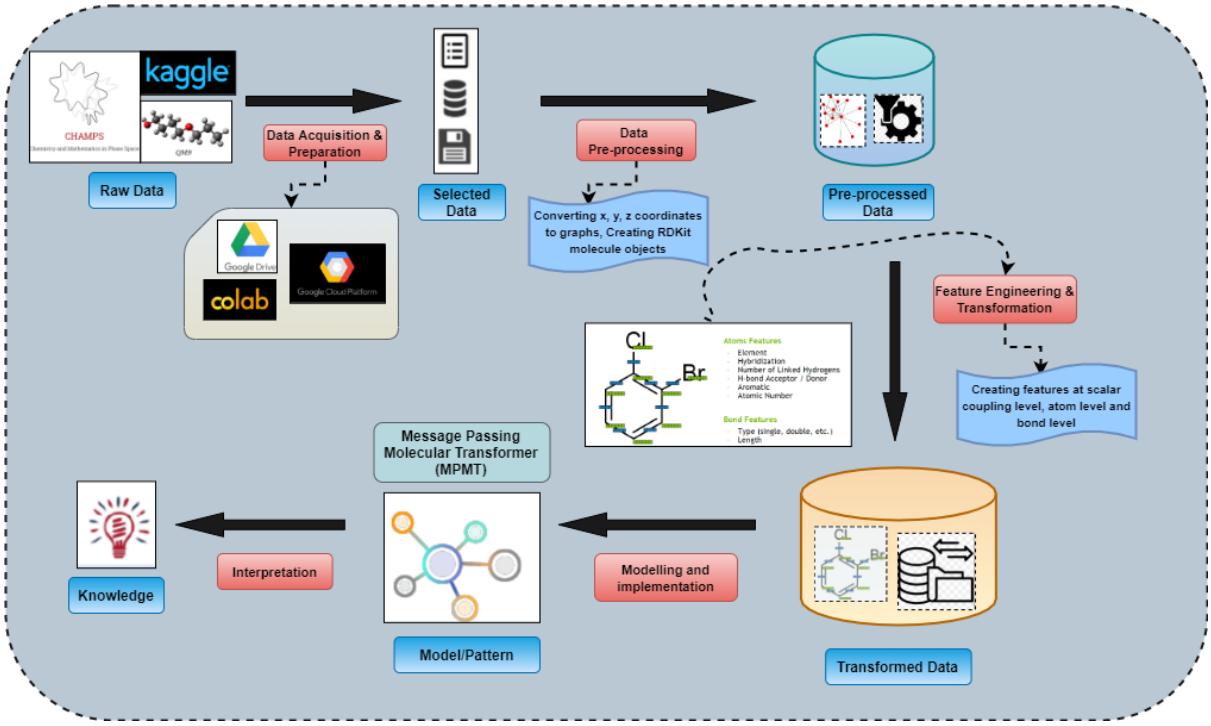


Figure 1: Research Methodology for the Study

3.1 Data Acquisition

Data acquisition is the first and most important step before commencing any research. For our research, the data chosen to demonstrate the proposed MPMT (Message Passing Molecular Transformer) model is the CHAMPS dataset (CHAMPS Dataset-Kaggle). The CHAMPS dataset has been collected by a team of four institutions - Bristol University, Cardiff University, Imperial College and Leeds University under the sponsored CHAMPS (Chemistry and Mathematics in Phase Space) program. It is publicly available on kaggle for research purposes. The CHAMPS dataset is actually derived from the QM9 dataset (Ruddigkeit et al. (2012), Ramakrishnan et al. (2014)) (QM9 Dataset) by making some modifications. This dataset contains the *.xyz* format files which contains the structural information (x,y,z coordinates) about the molecules (recognized by their unique numeric ID). The coordinates are also provided in a csv file along with the molecule names and index of the atoms involved. There are 8 different coupling types in the dataset based on the type and number of atoms involved in the interaction. These are $1J_{HC}, 1J_{HN}, 2J_{HH}, 2J_{HC}, 2J_{HN}, 3J_{HH}, 3J_{HC}, 3J_{HH}$. The distribution of these 8 coupling types in the training and test data is shown in the figure 2. The target variable i.e *scalar coupling constant* for atoms

is present in the *train.csv* file. The description of both files is given in figures 3a and 3b. Additionally, there are 5 more csv files containing the information about different properties of the molecules. These files are: *dipole moments.csv, magnetic_shielding_tensors.csv, mulliken_charges.csv, potential_energy.csv.*

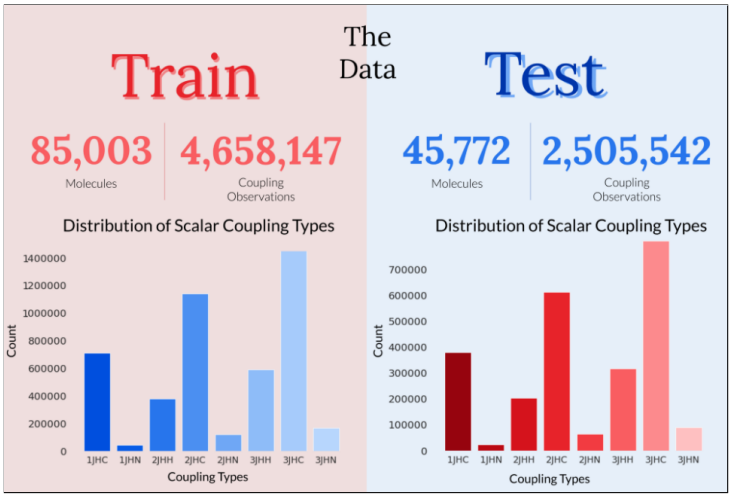


Figure 2: Distribution of different coupling types in the train and test files

Column_Name	Type	Description
id	Integer	Index of the rows (no significance as such)
molecule_name	String	Name of the molecule where the coupling constant originates
atom_index_0	Integer	Atom index of the first atom from the pair creating the coupling
atom_index_1	Integer	Atom index of the second atom from the pair creating the coupling
type	String	Type of coupling (8 unique values)
scalar_coupling_constant	Float	Value of the scalar coupling constant of the atomic pair

(a) Description of *train.csv*

Column_Name	Type	Description
molecule_name	String	Name of the molecule where the coupling constant originates
atom_index	Integer	Index of the atom
atom	String	Name of the atom
x	Float	x coordinate of the atom
y	Float	y coordinate of the atom
z	Float	z coordinate of the atom

(b) Description of *structures.csv*

Figure 3: Data description for CHAMPS dataset

3.2 Data Pre-processing

The data pre-processing was performed in two different stages. The following subsections describe each stage in detail.

- *Converting the Structural Information into Graphs*: This subsection describes how the structural information of the molecules (like x, y, z coordinates) was converted into graph structures for visualizing the relationship between the different atoms sharing a common coupling type.
- *Creating RDKit Molecule Objects*: This subsection describes how different molecule objects were created using the RDKit library in python and how these will be used in further pre-processing steps.

3.2.1 Converting the Structural Information into Graphs

The structural information of the molecules, i.e the x, y and z coordinates were first converted into graph objects and plotted to visualize the relationships between the atoms and the molecules. These are shown in figure 4.

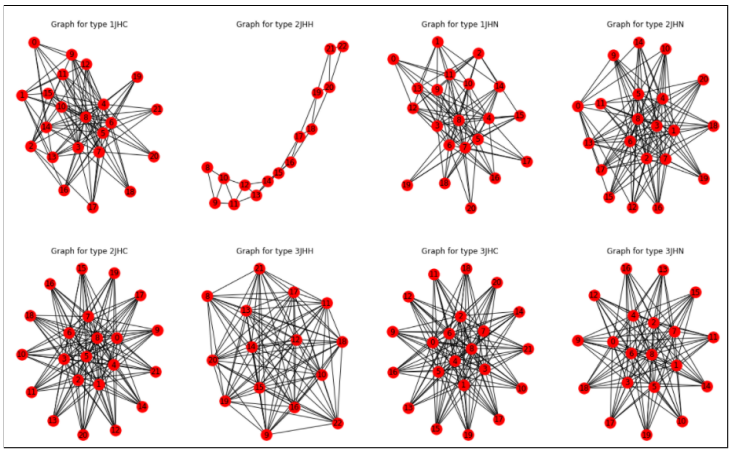


Figure 4: Graph representation of the data

3.2.2 Creating RDKit Molecule Objects

After the relationships between the atoms was visualized, the coordinates of the atoms were gathered and converted into molecule objects. The molecule level features were stored in the adjacency matrix. A GitHub repository created by Jan H. Jensen named *xyz2mol* was utilised to convert the xyz file into RDKit molecule objects. The function *get_molecules()* in the code file was designed to accomplish this. This function returned the following:

- The molecules IDs: A unique set of string identifying each molecule in the dataset.
- A set of molecule level features: The set of molecular features are returned as a pandas series.
- An array of the xyz coordinates: A numpy array of the x, y and z coordinates are returned.
- The graph distance matrix: The graph distance matrix contains the distances of every atom(nodes) from every other atom in the coupling type.

3.3 Feature Engineering and Transformation

The feature engineering is a major and most essential part of our research methodology since the performance of our model depends largely on the number and the importance of the features engineered from the existing data.

3.3.1 Creating Features at Scalar Coupling Level

The following series of steps were followed for creating the features at the scalar coupling level:

- First the coordinates of the atoms were merged into a pandas dataframe.
- Then the euclidean distances for all the atoms were calculated and appended into the above dataframe. Then we calculate the sum of inverse distances of SCC grouped at the atomic level.
- After that, the number of atoms (total and per coupling type) were added into the dataframe created above.
- The dihedral and the cosine angles using the dot product formula were then calculated and added to the dataframe along with the different edge and molecule level features.
- The scalar coupling contributions were then added to the train dataframe and then normalized by subtracting the mean from each value in the target column and then dividing the result by standard deviation of the column.

3.3.2 Creating Atom and Bond Level Features

Since there are two types of entities in a graph data structure i.e nodes and edges, there are two types of features in the data the graph is operating on, corresponding to each level - node and edge. The node (atom) level features include the element type, hybridization, number of linked hydrogen atoms, atomic number, aromatic, etc. On the other hand, the edge (bond) level features include bond length, bond type, etc. For the purpose of our study, we engineer some of these features and encode some of them using the one-hot encoding method. The final engineered atom and bond level features for our model is given in table 1. After this step, all the features engineered were stored into pandas data frames along with the distance matrices. The cosine angles for all the bonds were computed and were stored into data frames iteratively along with the atom index through a number of for loops.

Node (Atom) Level Features	**Edge (Bond) Level Features**
1. **Atom Type**: H, C, N, O, F (one-hot)	1. **Bond Type**(categorical): 1: single, 2: double, 3: triple, 4: aromatic (one-hot)
2. **Degree**: 1, 2, 3, 4, 5 (one-hot)	2. **Is Conjugated**: bool 0, 1
3. **Hybridization**: SP, SP2, SP3 (one-hot)	3. **Is in Ring**: bool 0, 1
4. **Is Aromatic** (bool): 0, 1	4. **Euclidean Distance**: float
5. **Formal Charge**: int	5. **Normalized Euclidean Distance**: float
6. **Atomic Number**: float	6. **Average Bond Length**: float
7. **Average Weight of Neighboring Atoms**: float	
8. **Donor**: bool 0, 1	
9. **Acceptor**: bool 0, 1	

Table 1: Node and Edge Level Features for Message Passing Molecular Transformer (MPMT)

3.4 Modelling

After analyzing the data, studying relationships between the atoms using graphs and engineering different features, it's now time to put all these together into our proposed model. The proposed model MPMT (Message Passing Molecular Transformer is a blend of two different architectures combined into one. It inherits the message passing elements from the Message Passing Neural Network (MPNN) first proposed by Gilmer et al. (2017). Gilmer used MPNN on the QM9 dataset, from which the CHAMPS dataset has been constructed. Although, the paper predicts bulk properties of the molecules and we are only interested in the edges,modifications can be made in the model so as to fit the requirement. The other half of our model is inspired by the revolutionary work of Vaswani et al. (2017) proposing the transformer architecture. The transformer architecture serves as the basic building block for our model. The only difference is that the encoder block in our model is made of 2 message passing layers which is followed by 3 multi-head attention layers, each layer followed by the layer normalization and finally followed by a feed-forward network (same as in transformer architecture).

3.5 Evaluation

After the model has been implemented, it is important to evaluate the results/predictions obtained in order to establish the validity of the proposed model. For our proposed MPMT model, the following evaluation metrices will be used:

- **Log Mean Absolute Error (LMAE)**: The log MAE is determined by calculating the logarithm of MAEs for each of the coupling types and then taking the average over all the types. The formula for the same is given as below. Here T is the total number of coupling types which in our dataset is seven. n_t is no. of observations of type t and y_i and $\hat{y}_i$ are the actual and the predicted values of the target variable i.e SCC.

$$LMAE = \frac{1}{T}\sum_{t=1}^{T}\log\left(\frac{1}{n_t}\sum_{i=1}^{n_t}|y_i - \hat{y}_i|\right) \tag{1}$$

- **Root Mean Squared Error (RMSE) Loss**: It is the square root of the mean of squared errors for each coupling type. There is a *contribs_rmse_loss()* function in the code which facilitates the

computation of this evaluation metric. It returns the sum of RMSEs for each scalar coupling contribution and SCC in a batch. It takes 2 different tensors of dimensions 2 and 3 which contain the predictions from the model and the true values of the coupling constants.

4 Design Specification

The design flow for the research is discussed in detail in the following subsequent subsections:

4.1 Multi-Head Attention and Message Passing Layers

As described in the section 3, our proposed model i.e Message Passing Molecular transformer (MPMT) is a blend of two completely different types of architectures - a Message Passing Neural Network and a Transformer. The MPMT contains 3 multi-head attention layers, with layer normalization after each layer, same as in the transformer encoder architecture and 2 message passing layers. Each layer in the multi head attention block computes attention using the features that we engineered in 5.3. The graphical representation of all the different layers and the overall architecture of MPMT is shown in the figure 6. The three attention layers are described below:

- *Layer 1 - Gaussian attention layer*: Uses Euclidean distance between the atoms to compute attention
- *Layer 2 - Graph distance attention layers*: Uses the embedding of the graph distance matrix (in other words, graph distances) to compute attention.
- *Layer 3 - Scaled Dot Product Attention*: This the same input layer as in the transformer encoder. According to Vaswani et al. (2017) It uses the dot product of the query with all the keys and divided each product by $sqrtd_k$ followed by a softmax function for obtaining the weights on the values. Here $\sqrt{d_k}$ is the dimension of keys and queries.

The two message passing layers are described below. The entire message passing algorithm is described in figure 5.

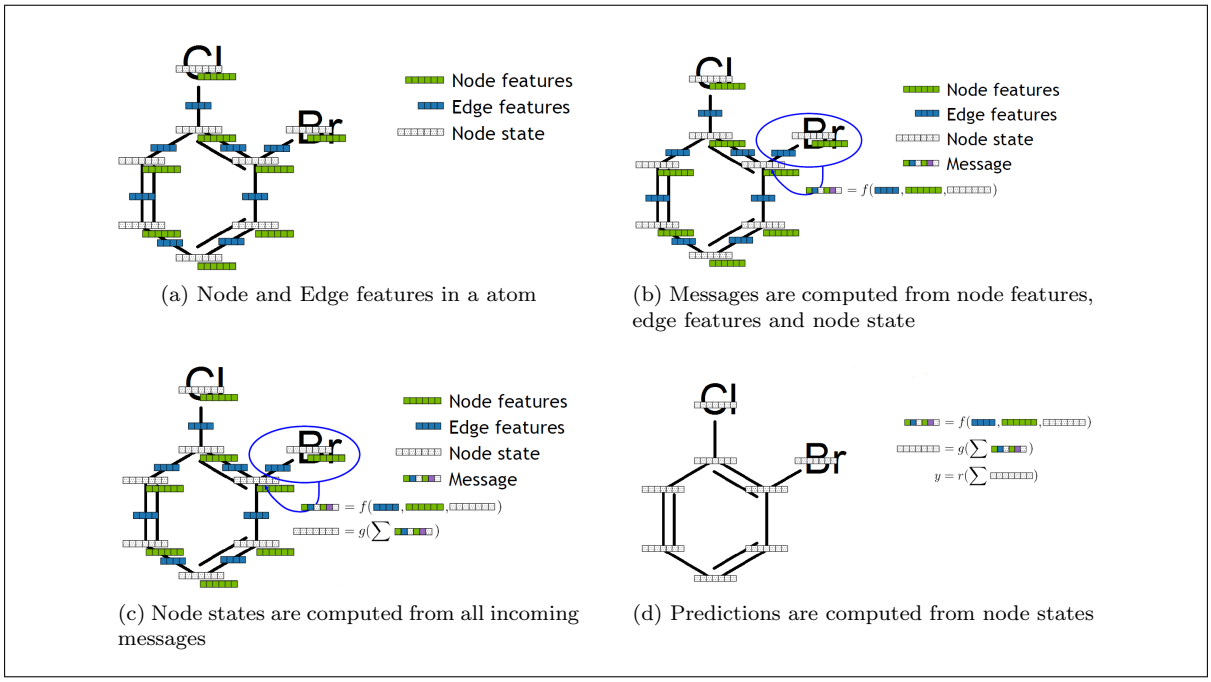
- *Bond Message Passing Layer*: Uses the bond features and cosine angles that were described in the previous subsection to compute the message for the next edge.
- *Scalar Coupling Message Passing Layer*: Uses the scalar coupling level features to compute the message for the next node.

4.2 MPMT Architecture

The major changes introduced in the existing MPNN architecture that allowed the successful implementation of MPMT model are summarized below. The final model architecture is shown in the figure 6.

- Introduction of separate MP functions for the atomic pairs.
- Engineering various node, edge and molecule level features. Specifically, dihedral angles for coupling type 3J and cosine angles for type 2J).
- Addition of attention mechanism on the incoming message functions on the basis of cosine angles.
- Replacing the *set2set()* function with an attention layer based on the Euclidean distance between the atoms.
- Before making the final predictions, the four coupling contribution terms (*fc, sd, pso and dso*) were computed separately first. Then all of them were added to the residual term for predicting the final SCC. These predictions for these contributions were added to the loss function.

(**Note**: The SCC in the *train.csv* file is sum of four terms. These are Fermi-Contact (fc), Spin-dipolar (sd), Para-magnetic spin-orbit (pso) and Diamagnetic spin-orbit (dso). These four contribution terms can be found in the *scalar_coupling_contributions.csv* file mentioned in the data description in section 3.1.)



(a) Node and Edge features in a atom

(b) Messages are computed from node features, edge features and node state

(c) Node states are computed from all incoming messages

(d) Predictions are computed from node states

Figure 5: Visual representation of Message Passing Algorithm

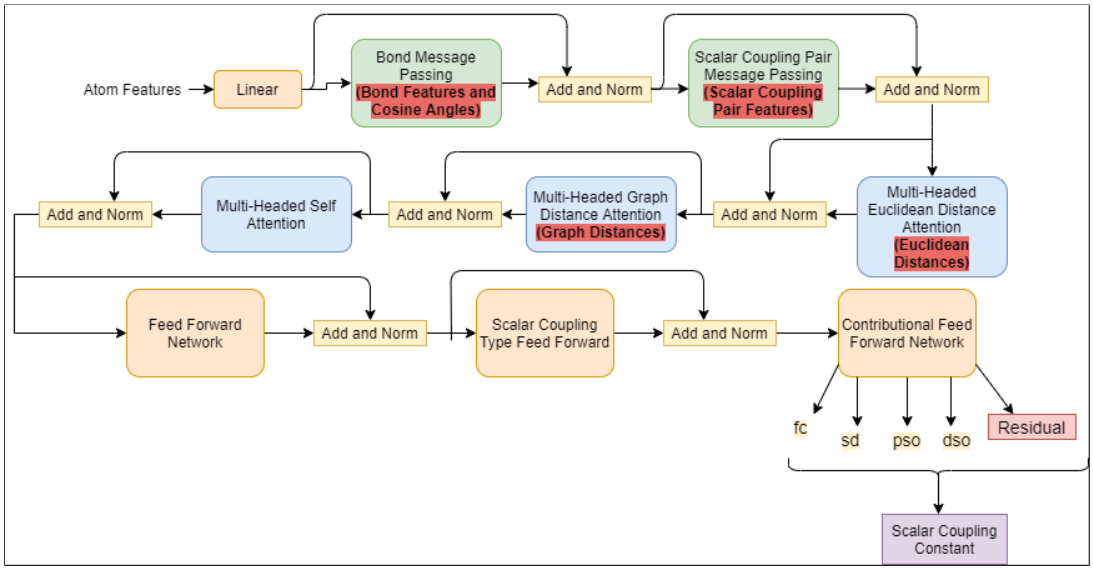


Figure 6: Representation of MPMT architecture

5 Implementation

5.1 Softwares and Technologies Used

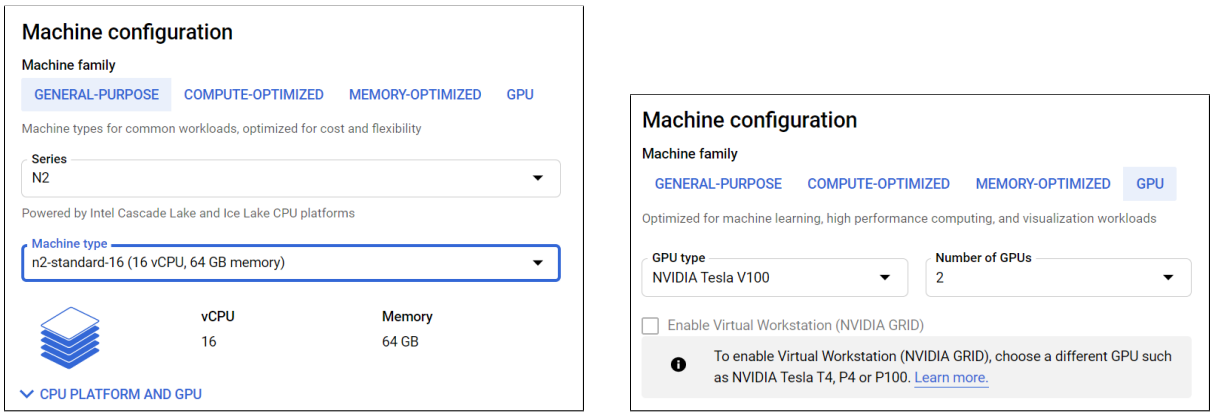
The following specifications for the softwares/libraries were used to produce the outputs and results.

- **Cloud Platform for Model Training**: GCP (Google Cloud Platform)
- **Programming Language**: Python
- **Database management**: Google Drive
- **IDE**: Google Colab, Jupyter Lab, VS Code
- **Python Libraries/Modules**:
  - *Pre-Processing*: Numpy, Pandas
  - *Feature Engineering*: RDKit, DeepChem, xyz2mol (a custom GitHub repository made by Jan H. Jensen and made publicly available. It is based on the work of Kim and Kim (2015))
  - *Modelling and Evaluation*: Sklearn, PyTorch, Fastai (for speeding up the training process), Fcnet, Layernorm

5.2 Data Preparation on Google Cloud Platform (GCP)

The preparation of the data was an important aspect of this research. Before starting any pre-processing, the data was uploaded to google drive. Then the Google Cloud Platform was set up by creating a VM instance as described in figure 7. The specifications selected for the instance are specified below. The other instance properties like SSH keys, security, etc. were chosen for their default values.

- OS: Ubuntu 18.04LTS
- *CPU*: No. of cores: 16, Memory: 64GB
- *GPU*: 2 NVIDIA Tesla V100



(a) CPU specifications

(b) GPU specifications

Figure 7: Configurations for GPU and CPU for the virutal instance on GCP

5.3 Feature Engineering

The feature engineering was the major part of the implementation of our proposed model. The more features we can engineer, the better will be predictions. This is because, the scalar coupling is an inherent property of the protons (which are present inside the nucleus of he atom) and despite the huge success of the quantum mechanics in computing the scalar couplings over the years, we still do not completely understand what specific properties of the element really affect the magnetic interaction between them. The different types of features engineered are described below:

- **Coupling Level Features**: These features are basically those features which can directly influence the magnetic interaction or in other words the value of scalar coupling constant between the atoms. These features include the Euclidean distances between the atoms, inverse distances of SCC grouped at the atomic level, harmonic mean of inverse distances, dihedral and cosine angles and the scalar coupling edge and molecule level features.
  - First the coordinates of the atoms were merged into a pandas dataframe. Then the euclidean distances for all the atoms were calculated and appended into the dataframe. Then we calculate the sum of inverse distances of SCC grouped at the atomic level.
  - We then calculate the dihedral and the cosine angles using the dot product formula and add these features into the dataframe.
  - We then add the edge and molecule level features to the dataframe. These features were returned by the *get_molecules()* function. The scalar coupling contributions were then added to the train dataframe and then normalized by subtracting the mean from each value in the target column and then dividing the result by standard deviation of the column.
- **Node and Edge Level Features**: Node and edge level features are the features corresponding specifically to the atoms and bonds. In other words, these features only affect the particular atom or bond they are involved with. Node or atom level features include atomic type, degree, hybridization, charge, average weight of neighbouring atom, or whether the atom is a donor or an acceptor. Edge or bond level features include bond type, bond length, Euclidean distance (normalized), whether the atom is part of ring structure, etc.

5.4 Gradient Clipping

For the training of the final MPMT model, the gradient clipping technique was adopted in order to prevent the overshooting gradients. Under this method, we changed the differential of the error before the start of back propagation through the network and then using it for updating the weights.

5.5 Model Training

The training for the final model was performed using the fastai library. The use of this particular method was adopted because the size of the model was significantly large and the fastai library facilitates the faster training of models with just a few lines code. The model training was performed on the Google Cloud Platform (GCP). The final MPMT model has been trained for 100 epochs using the one cycle learning rate method. The model had the maximum learning rate of 5×10^{-4} and a decay rate of 10^{-2}. The final training was done using the Google Cloud Platform using the distributed training method on 2 NVIDIA Tesla V100 GPUs. Total training time was roughly around 14 hours for the final model.

6 Evaluation

The following experiments were performed and their results were evaluated based on the evaluation metrices described in section 3.5. The results of all the experiments have been discussed in detail below.

6.1 Experiment 1

The aim of this experiment is to build a message passing neural network (MPNN) which can predict the magnetic interactions between the atoms (i.e the values of scalar coupling constant).

6.1.1 Modelling

This experiment is based on and is inspired by the work done by Withnall et al. (2020) for predicting molecular property by using edge message passing network. The different layers in the model are as follows:

- Message Passer
- Aggregator
- Node Update Layer

- Edge Update Layer
- Output Layer (Edge Regressor)
- Message Passing Layer

6.1.2 Results

The model was able to achieve the log MAE value of -2.19 which translates to an error in the scalar coupling value of about 0.11 Hz. The 4 best predictions of the coupling types is shown in the figure 9. The loss curve is shown in the figure 8 and the plots of predicted vs the actual target values is shown in the figure 9. In practice, it means that MPNN could be used as a viable alternative to the different machine learning approaches that we discussed for predicting the various molecular properties. Various different variations of the MPNN architecture, can also be employed, depending upon the prediction task to be performed.

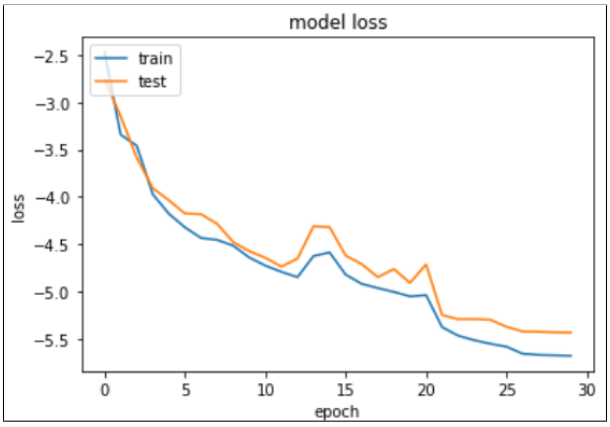
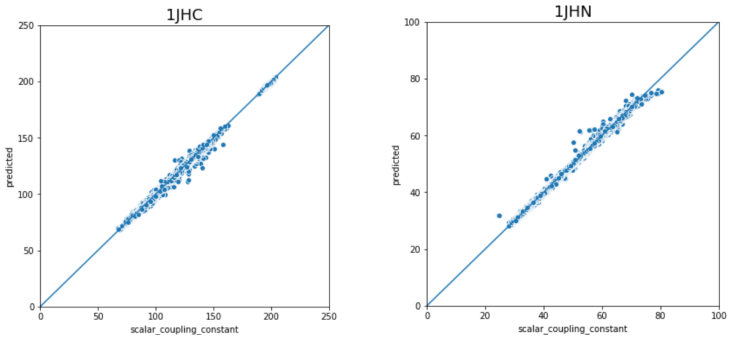


Figure 8: Loss curve for training and test data



(a) 1 JHC (b) 1 JHN

Figure 9: Predicted vs Target Values for Coupling Constant Predictions for MPNN model

6.2 Experiment 2

The aim of this experiment is to build the proposed Message Passing Molecular Transformer network for the predicting the values of scalar coupling constants between the atoms.

6.2.1 Modelling

The model consists of the 3 multi-head attention layers and two message passing layers. In addition, it uses layer normalization after each sub layer. These 5 layers are followed by 3 feed forward networks. The final feed forward network computes the coupling contributions from all the four terms separately and stores them in the loss function. These four terms and the residual is used for predicting the SCC for the atoms in the *train.csv* file.

6.2.2 Results

The final MPMT model with a learning rate of 5×10^{-4} and the weight decay of 10^{-2} was able to achieve the log MAE value of -2.873 on a 8-fold MPNN architecture. The loss curve for the model is shown in the figure 10. The plot for the predicted vs target values for the two coupling types with best log MAE values is shown in the figure 11. In practice, it means that our proposed message passing molecular transformer (MPMT) model has outperformed the state-of-the-art work on the CHAMPS dataset (Jian et al. (2020)) which reported MAE of 0.096 Hz, achieving an overall MAE of 0.056 Hz. This means that our study can be used to set a new standard in molecular property prediction domain for the benchmarked CHAMPS dataset.

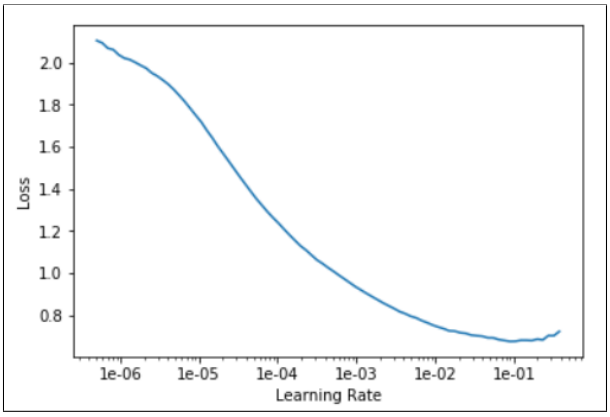
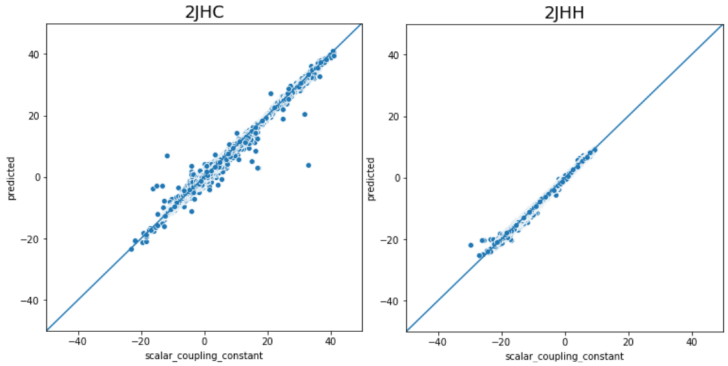


Figure 10: Loss Curve for the MPMT model



(a) 2 JHC (b) 2 JHH

Figure 11: Predictions for the MPMT model for 2JHC and 3JHN coupling types

6.3 Discussion

This study shows that the Message Passing Molecular Transformer model was able to outperform the state-of-the-art results obtained by Jian et al. (2020) for the CHAMPS dataset. It was able to achieve an overall LMAE score of -2.873 which corresponds to the mean absolute error value of 0.056 Hz compared to the MAE value of 0.096 Hz reported by Jian et al. (2020). For the MPNN model, we were able to achieve an overall LMAE of -2.19 corresponding to absolute error of 0.111 Hz. Through the plot shown in figures 8, we can see that the loss decreases smoothly with the increase in the learning rate which can be attributed to the choice for our weight decay parameter. From the predicted vs target variable plot shown in figure 11, it can be seen that there are few values for the 2 JHC coupling type having large RMSE. This may be because of the fact that we have created a single model for all the coupling types. Creating different small graph networks for each coupling type could resolve this issue. Our proposed architecture was also able to outperform the machine learning based gradient boosting algorithm frameworks employed by Shibata and Kaneko (2021) and Zhang, Deng and Jia (2020) which reported an overall MAE of 1.23

Hz and 4.34 Hz respectively, for the prediction of scalar coupling constant. This means that our research can set a new standard for the state-of-the-art for the benchmarked CHAMPS dataset.

This study can contribute to the field of the molecular property prediction by significantly reducing the training times for SCC predictions as compared to the quantum computations. Our research has the potential to consequentially contribute towards the development of new drugs by significantly reducing the research costs. It can also aid in studying the protein structures in different organisms and also improving the current NMR techniques by enhancing our understanding of the molecular structure. most importantly adding to our knowledge of different physical and biological systems in nature.

7 Conclusion and Future Work

In this study, our aim was to design and investigate a graph neural network for predicting the magnetic interaction between the atoms in a molecule. We designed a Message Passing Neural Network (MPNN) and a novel Message Passing Molecular Transformer (MPMT) for determining the coupling constants for atoms in the CHAMPS dataset. A number of node, edge and molecule level features were engineered. Each coupling type in the dataset was studied individually first through a series of exploratory analysis using graph structures. After a thorough critical review of the current state-of-the-art methods, a novel Message Passing Molecular Transformer (MPMT) architecture was proposed, designed and implemented on the benchmarked CHAMPS dataset (derived from the QM9 dataset). The proposed architecture was evaluated on the basis of log MAE score averaged over the 8 coupling types. It was able to achieve an overall log MAE of -2.873 which corresponds to a mean absolute error of about 0.056 Hz. The state-of-the-art result for the CHAMPS dataset is 0.096 Hz which was produced by Jian et al. (2020).

There is plenty of room for the future research purposes in the domain. Although, our proposed model produced very good predictions, it was pretty big to be trained on a single GPU. Therefore, we utilized the distributed training feature in pytorch and trained the model on two Tesla V100 GPUs on the google cloud platform (GCP). A smaller model could be designed, without having the need to engineer features, using just the message passing network (as described by Withnall et al. (2020)) with a few modifications, without loosing the efficiency of the model. Another direction for the future research where some progress can be made is the feature engineering part of the methodology. If we can rank the engineered features according to their importance, better prediction models can be constructed using those descriptors. This issue was also highlighted by Shibata and Kaneko (2021) in their study. We have built a single neural network for all the coupling types, that's why MAE values for some coupling types is better than the others. This limitation could be improved by building different graph neural networks for each of the coupling types based on their distribution and graph structure.

If deployed on a larger scale with enough computing resources, our proposed architecture could prove more valuable to the researchers and scientists. With some modifications, it could even be used as a black box by scientists for predicting not just the scalar coupling constant, but many other molecular properties like drug affinity, toxicity, etc. as well.

References

Battaglia, P. W., Pascanu, R., Lai, M., Rezende, D. and Kavukcuoglu, K. (2016). Interaction networks for learning about objects, relations and physics, *arXiv preprint arXiv:1612.00222* .

Coley, C. W., Barzilay, R., Green, W. H., Jaakkola, T. S. and Jensen, K. F. (2017). Convolutional embedding of attributed molecular graphs for physical property prediction, *Journal of chemical information and modeling* **57**(8): 1757–1772.

Fout, A. M. (2017). *Protein interface prediction using graph convolutional networks*, PhD thesis, Colorado State University.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. and Dahl, G. E. (2017). Neural message passing for quantum chemistry, *International conference on machine learning*, PMLR, pp. 1263–1272.

Gryff-Keller, A. and Szczeciński, P. (2016). An efficient dft method of predicting the one-, two-and three-bond indirect spin–spin coupling constants involving a fluorine nucleus in fluoroalkanes, *RSC advances* **6**(86): 82783–82792.

Hamilton, W. L., Ying, R. and Leskovec, J. (2017). Inductive representation learning on large graphs, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 1025–1035.

Hechtlinger, Y., Chakravarti, P. and Qin, J. (2017). A generalization of convolutional neural networks to graph-structured data, *arXiv preprint arXiv:1704.08165* .

Jian, C., Cheng, X., Zhang, J. and Wang, L. (2020). Scalar coupling constant prediction using graph embedding local attention encoder, *IEEE Access* **8**: 171100–171111.

Jørgensen, P. B., Jacobsen, K. W. and Schmidt, M. N. (2018). Neural message passing with edge updates for predicting properties of molecules and materials, *arXiv preprint arXiv:1806.03146* .

Kearnes, S., McCloskey, K., Berndl, M., Pande, V. and Riley, P. (2016). Molecular graph convolutions: moving beyond fingerprints, *Journal of computer-aided molecular design* **30**(8): 595–608.

Kim, Y. and Kim, W. Y. (2015). Universal structure conversion method for organic molecules: from atomic connectivity to three-dimensional geometry, *Bulletin of the Korean Chemical Society* **36**(7): 1769–1777.

Komorovsky, S., Jakubowska, K., Swider, P., Repisky, M. and Jaszunski, M. (2020). Nmr spin–spin coupling constants derived from relativistic four-component dft theory—analysis and visualization, *The Journal of Physical Chemistry A* **124**(25): 5157–5169.

Pu, Q., Li, Y., Zhang, H., Yao, H., Zhang, B., Hou, B., Li, L., Zhao, Y. and Zhao, L. (2019). Screen efficiency comparisons of decision tree and neural network algorithms in machine learning assisted drug design, *Science China Chemistry* **62**(4): 506–514.

Ramakrishnan, R., Dral, P. O., Rupp, M. and Von Lilienfeld, O. A. (2014). Quantum chemistry structures and properties of 134 kilo molecules, *Scientific data* **1**(1): 1–7.

Ruddigkeit, L., Van Deursen, R., Blum, L. C. and Reymond, J.-L. (2012). Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17, *Journal of chemical information and modeling* **52**(11): 2864–2875.

Shibata, K. and Kaneko, H. (2021). Prediction of spin–spin coupling constants with machine learning in nmr, *Analytical Science Advances* .

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I. (2017). Attention is all you need, *Advances in neural information processing systems*, pp. 5998–6008.

Withnall, M., Lindelöf, E., Engkvist, O. and Chen, H. (2020). Building attention and edge message passing neural networks for bioactivity and physical–chemical property prediction, *Journal of cheminformatics* **12**(1): 1–18.

Xiong, Z., Wang, D., Liu, X., Zhong, F., Wan, X., Li, X., Li, Z., Luo, X., Chen, K., Jiang, H. et al. (2019). Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism, *Journal of medicinal chemistry* **63**(16): 8749–8760.

Yun, S., Jeong, M., Kim, R., Kang, J. and Kim, H. J. (2019). Graph transformer networks, *Advances in Neural Information Processing Systems* **32**: 11983–11993.

Zhang, X.-x., Deng, T. and Jia, G.-z. (2020). Nuclear spin-spin coupling constants prediction based on xgboost and lightgbm algorithms, *Molecular Physics* **118**(14): e1696478.

Zhang, Z., Cui, P. and Zhu, W. (2020). Deep learning on graphs: A survey, *IEEE Transactions on Knowledge and Data Engineering* .