Exploring Graph Neural Networks for Molecular Activity Prediction

Exploring Graph Neural Networks for Molecular Activity Prediction

Guangyu Song

A Thesis in the Field of Software Engineering

for the Degree of Master of Liberal Arts

Harvard University

May 2024

Abstract

Graph Neural Networks (GNNs) have emerged as a powerful class of machine learning techniques capable of processing graph-structured data, showing immense promise in molecular property prediction. This thesis compares the performance of two specific GNNs—Graph Attention Networks (GATs) and Attentive FP models with Graph Convolutional Networks (GCNs) in predicting the biological activities of protein targets across several protein families. Our findings indicate that while GCNs are highly effective for molecular property prediction, GATs and Attentive FP models also offer competitive performance, with GATs showing particular promise for enzymes and transporters. The experiments suggest that the choice of model should be tailored to specific families of target proteins, highlighting the need to consider the particular protein family when selecting GNNs for predictive modeling in drug discovery.

Dedication

I dedicate this to those who spend their lives in the pursuit of knowledge.

Acknowledgements

I would like to thank my thesis director, Professor Hongming Wang for her guidance and support throughout this journey. Her patience and encouragement have been invaluable to me.

Contents

Table of Contents

List of Figures

List of Tables

Chapter I.

Introduction

1.1. Background

Deep learning have emerged as a promising method to aid the discovery of new medicine and reduce the time and cost associated with experimental screening (Gawehn et al., 2016). With the advent of these techniques, more sophisticated computational methods for molecular activity prediction have been developed (LeCun et al., 2015; Chen et al., 2018; Zhavoronkov et al., 2019; Stokes et al., 2020; Jiménez-Luna et al., 2020). Graph neural networks (GNNs) have shown great promise in modeling the relationships between molecular structures and their properties Wu et al. (2021).

GNNs are adept at processing graph-structured data, which makes them ideal for modeling the complex relationships between atoms and bonds in a compound (Gilmer et al., 2017). GNNs learn representations of molecular structures directly from input data, addressing limitations of traditional models like Quantitative Structure-Activity Relationships (QSAR) and Quantitative Structure-Property Relationships (QSPR), thereby enhancing predictive performance (Gawehn et al., 2016). Among

various GNN architectures, GATs are notable for their effective depiction of complex node relationships within graphs through the use of attention mechanisms, which allow them to adaptively weight the influence of neighboring nodes (Veličković et al., 2018). This adaptability makes GATs particularly suited for handling complex molecular structures and predicting a broad range of properties, such as molecular activity (Wu et al., 2021). Recent research has shown GATs' capabilities in predicting molecular properties like solubility (Lee et al., 2023) and toxicity (Chen et al., 2021; Cremer et al., 2023), underlining their significance in advancing molecular activity prediction.

Molecules can be effectively represented as graphs, with atoms as nodes and bonds as edges (Gilmer et al., 2017). Additional details, such as atom types, bond types, and other atom-level features, can be incorporated as attributes of the nodes and edges. This graph-based representation facilitates the direct encoding of a molecule's topological structure, which plays a crucial role in determining its properties and molecular activities (Gawehn et al., 2016). Once the molecular graph is constructed, GNNs can be employed to process and analyze the graph-structured data.

A key component of GNNs is the message-passing framework, which allows for the efficient and systematic aggregation of information from neighboring nodes. Node representations are iteratively updated and optimized through a series of message-passing steps that combine the features of neighboring nodes and edge attributes. This process enables the GNN to capture both local and nonlocal information within the molecular graph, ultimately resulting in a fixed-size representation that can be

used for various property prediction tasks (Gilmer et al., 2017).

One of the most widely used GNN architectures for molecular property prediction is the Graph Convolutional Network (GCN) (Kipf & Welling, 2017). GCNs employ a convolutional layer to update node representations by aggregating information from neighboring nodes through a simple averaging operation. Sakai et al. (2021) utilized GCNs to predict pharmacological activities based on chemical structures with graph convolutional neural networks. Despite their success in various molecular property prediction tasks, GCNs also have several limitations. For example, the simple averaging operation can lead to information loss, as it does not consider the relative importance of different neighboring nodes (Kipf and Welling, 2017). Furthermore, GCNs might struggle to capture complex relationships between nodes, as they lack a mechanism for adaptively weighting the contributions of neighboring nodes.

These limitations of GCNs have motivated the development of more advanced GNN architectures, such as Graph Attention Networks (GATs), which aim to address these shortcomings and improve the performance of molecular property prediction tasks (Veličković et al., 2018). By incorporating attention mechanisms, GATs are able to adaptively weight the contributions of neighboring nodes, allowing for a more flexible and robust approach to information aggregation (Veličković et al., 2018). As a result, GATs have the potential to outperform traditional GCNs and other GNN architectures in predicting molecular activity and other molecular properties (Wu et al., 2021). In addition to GATs, several other advanced GNN architectures have emerged

in recent years, further contributing to the ongoing advancements in the field of molecular property prediction. These include Graph Isomorphism Networks (GINs), which aim to capture the structural information of molecular graphs more effectively by considering graph isomorphism, and Message Passing Neural Networks (MPNNs), which provide a general framework for constructing a wide variety of message-passing algorithms for GNNs (Gilmer et al., 2017). Each of these architectures offers unique advantages and addresses different aspects of the limitations found in earlier GNN models.

The growing body of research on GNNs for molecular property prediction underscores the potential of these techniques to revolutionize drug discovery and related fields. While GCNs and GATs have demonstrated significant success in various prediction tasks, ongoing research continues to explore and develop new GNN architectures that can further enhance the performance and interpretability of these models. By understanding the strengths and weaknesses of different GNN architectures, researchers can better select and tailor GNN models to address specific challenges in molecular property prediction tasks (Stokes et al., 2020). The continued advancement of GNNs and their application to molecular property prediction has the potential to significantly impact the efficiency of the drug discovery process, paving the way for more effective and targeted therapeutic interventions (Chen et al., 2020).

GATs have emerged as a powerful GNN architecture for molecular property prediction, largely due to their incorporation of attention mechanisms (Veličković

et al., 2018). Unlike GCNs, which aggregate information from a one-hop neighborhood uniformly, GATs employ attention mechanisms to dynamically assign weights to neighboring nodes during aggregation, based on a computed compatibility score between nodes. This typically results in a more flexible and robust approach than GCN's reliance on simple averaging operations (Kipf & Welling, 2017). By learning to assign different levels of importance to neighboring nodes, GATs can better capture the complex relationships between atoms and bonds within a molecule, ultimately leading to more accurate property predictions (Wu et al., 2021). The advantages of attention mechanisms in GATs can be attributed to their ability to focus on the most relevant parts of a molecular graph while filtering out less important information (Veličković et al., 2018). This selective focus enables GATs to effectively capture both local and global contexts within the molecular graph, thereby enhancing their performance in predicting various molecular properties (Wu et al., 2021). Additionally, the attention mechanism in GATs also provides a degree of interpretability, as it allows researchers to visualize the importance of different parts of the molecular graph in the prediction process (Veličković et al., 2018). Several studies have already demonstrated the potential of GATs for predicting molecular properties. For example, in Lusci et al. (2013), GATs were applied to predict the solubility of compounds, outperforming traditional GNNs and other machine learning methods in terms of prediction accuracy. Another study used GATs to predict the binding affinity of small molecules to protein targets, showcasing their ability to effectively model complex molecular interactions

(Chen et al., 2018). These studies, among others, highlight the effectiveness of GATs in various molecular property prediction tasks and suggest their potential for advancing the state of the art in molecular activity prediction (Stokes et al., 2020). Building on this concept, Attentive FP (Xiong et al., 2019) emerges as another powerful GNN model that uses Recurrent Neural Networks (RNNs) to aggregate the structural information encoded in the graph from nearby to distant nodes. It adopts graph attention mechanisms at both the atomic and molecular scales to discern local and nonlocal chemical properties. Attentive FP has been shown to be very successful for molecular toxicity prediction tasks (Ketkar et al., 2023).

1.2. Motivation

Chemicals are the building blocks of the world, and chemical compounds constitute all living things. By observing the properties of chemical compounds and their interactions, we understand the world and create drugs to achieve desired effects. Traditionally, drug discovery has been a manual process, but with technological advancements, we have accelerated this process. Although the advent of AI has automated much of it, there is still room to further our understanding.

While leveraging GNNs for predictive tasks in drug discovery has become commonplace, the industry often relies on a narrow selection of models. Typically, once these models are trained, they are frequently used indiscriminately for various tasks. This approach may not yield the best results as the most optimal model for

a particular task does not automatically mean it will be equally effective for other tasks. Thus, it is crucial to investigate diverse model architectures in order to enhance performance. By doing so, we can potentially uncover more efficient methods for drug discovery, tailoring model selection to the unique challenges of each task, and thereby pushing the boundaries of what is possible with machine learning in this domain.

1.3. Problem Domain

Molecular property prediction, a crucial task in drug discovery, demands accurate methods to infer properties from molecular structures. This domain encompasses both categorical (e.g., toxicity) and continuous (e.g., solubility) property predictions, posing unique challenges. The effectiveness of predictions depends on the chosen representation method, such as Simplified Molecular Input Line Entry System (SMILES) or fixed representations like fingerprints and structure keys.

Molecular property prediction falls under the supervised learning category, where a model is trained on a dataset consisting of molecules with known properties. Once trained, the model is able to make predictions on the properties of new, unseen molecules. To evaluate the model's performance, a separate test set is used, which includes molecules with known properties. For classification tasks, metrics like accuracy, precision, recall, and F1 score are commonly employed. On the other hand, mean squared error (MSE), root mean squared error (RMSE), and the coefficient of determination (R^2), and area under the curve (AUC) are frequently used for

regression tasks.

1.4. Thesis Outline

The main goal of this thesis is to explore the prediction capabilities of GATs and Attentive FP networks in molecular property prediction. We compared their performance with GCNs on prediction tasks. Computational experiments were conducted using the BACE dataset, benchmark datasets from earlier papers, and a manually filtered dataset to demonstrate its effectiveness.

This thesis begins with the Introduction (Chapter I), which introduces the topic and sets the stage for the subsequent chapters. A critical review of the existing literature on GNNs and their applications is provided in the Graph Neural Networks section (Chapter II). The Methodology chapter (Chapter III) outlines the specific methodologies and experimental procedures employed. The findings of the research are presented in the Results chapter (Chapter IV), followed by the Discussion chapter (Chapter V), which interprets these findings and discusses their implications. The thesis concludes with the Conclusion (Chapter VI), summarizing the study's key insights and suggesting directions for future research.

Chapter II.

Graph Neural Networks

In the realm of neural networks, GNNs have marked a significant breakthrough with their specialized approach to graph-structured data. The foundational concept of GNNs dates back to the late 1980s, but it was not until the 1990s that these networks began to gain traction within the research community. The early work by Baldi & Chauvin (1996) laid the groundwork for what would eventually evolve into modern GNNs. In their seminal paper, they proposed a hybrid architecture combining elements of Hidden Markov Models (HMMs) and Neural Networks, addressing some of the limitations inherent in HMMs, such as control of model structure and complexity. This approach demonstrated significant improvements in efficiency, as evidenced by their construction of a model for the immunoglobulin protein family with fewer parameters than previous HMMs.

The term "Graph Neural Network" itself was coined later by Gori et al. (2005), conceptualized as an extension of recursive neural networks capable of directly processing graph data. This marked a pivotal moment in the history of GNNs, laying the foundation for their widespread application in various fields. The comprehensive

framework for GNNs, encompassing both graph and node-focused applications, was further refined and formally introduced by Scarselli et al. (2009). This work unified the diverse applications of GNNs into a common framework, setting the stage for the rapid development and adoption of GNNs in subsequent years.

Since then, GNNs have have proven to be valuable in various domains such as social media analysis, recommendation systems, drug discovery, natural language processing, and computer vision. The flexibility and applicability of GNNs make them an indispensable resource for modern machine learning methods, providing innovative responses to complex problems that can be framed within the graph-based approach.

A GNN typically consists of three types of layers, each serving a distinct purpose in processing graph-structured data. These layers are designed to capture the inherent structure of the graph and the relationships between its nodes, enabling the network to learn and predict based on this information. The three types of layers in a GNN are as follows:

(i) Permutation-equivariant layers: These layers map a graph to an updated representation of the same graph. They are commonly referred to as message passing layers, which propagate messages between nodes based on their neighborhood information.

(ii) Local pooling layers: These layers aggregate the node representations into a graph representation, helping to reduce the dimensions of the data by combining information from the nodes.

(iii) Global pooling layers: Also known as readout layers, these layers consolidate information from every node in the graph to provide a fixed-size representation of the entire graph.

2.1. Graph Neural Networks and Molecular Activity Prediction

Molecular activity prediction has become a popular application of GNNs in recent times. Traditional methods, such as QSAR and QSPR, primarily relied on descriptor-based models for molecular activity prediction. These approaches characterized molecules using a predefined set of descriptors and then mapped these to the properties of interest using linear or nonlinear models (Shayanfar et al., 2010; Perkins et al., 2003). GNNs, however, have shifted this paradigm by enabling end-to-end learning of molecular properties. This development represents a significant move from relying on expert intuition for feature engineering to allowing algorithms to automatically learn and discern underlying representations directly from the data. GNNs have demonstrated efficacy in predicting a variety of molecular properties, including solubility (Lee et al., 2023), toxicity (Chen et al., 2021; Cremer et al., 2023), and drug-likeness (Sun et al., 2022).

2.2. Fundamentals of GCNs

GCNs have emerged as a powerful tool in deep learning, particularly for tasks where data is naturally structured as graphs. Originating from advancements in CNNs

(Lecun et al., 1998) for grid-like data, such as images, GCNs adapt the concept of convolutions to graph-structured data.

The fundamental principle behind GCNs is message passing, which aggregates feature information from a node's neighbors in a graph. This process enables nodes to capture local graph structures and feature distributions. The aggregation function, typically a weighted sum, is crucial as it determines how neighbor information is combined at each node.

In a typical GCN layer, each node's features are updated by aggregating its own features with those of its neighbors, followed by a nonlinear transformation. The weights in the aggregation function are learned during training, allowing the network to adapt to specific patterns in the graph data.

GCNs have been successfully applied in various domains, including social network analysis (Lin, 2020), recommendation systems (Wu et al., 2019; Zhang et al., 2020), and bioinformatics Sun et al. (2020); Zhang et al. (2021). In the context of pharmacological activity prediction, GCNs leverage the structural information of chemical compounds represented as molecular graphs, where nodes represent atoms and edges represent bonds. Sakai et al. (2021) have attempted to use GCNs for predicting compound activities, highlighting the importance of graph-structural features and the effectiveness of GCNs in capturing intricate details crucial to pharmacology.

The core functionality of a GCN can be defined as:

$$h_{v_i}^{(l+1)} = \sigma\left(\sum_j \frac{1}{c_{ij}} h_{v_j}^{(l)} W^{(l)}\right),$$

where $h_{v_i}^{(l+1)}$ denotes the feature vector of node v_i at layer $(l+1)$, and σ is a nonlinear activation function. The term $\sum_j \frac{1}{c_{ij}} h_{v_j}^{(l)} W^{(l)}$ represents the aggregation of features from neighboring nodes (v_j) of node v_i, weighted by the layer-specific weight matrix $W^{(l)}$ and normalized by c_{ij}, which is a normalization constant that depends on the degree of the nodes v_i and v_j.

2.3. Emergence of Advanced GNN Architectures

While early GNNs like GCNs marked a significant advancement, they had limitations, such as a simplistic averaging operation that could lead to information loss (Kipf & Welling, 2017). This paved the way for more sophisticated architectures like GATs (Veličković et al., 2018) and Attentive FP (Xiong et al., 2019) networks, which incorporate attention mechanisms to improve predictive accuracy and interpretability.

Yang et al. (2019) introduced Directed Message Passing Framework (D-MPNN) for molecular property prediction, a model that uses a hybrid representation combining convolutions and descriptors. D-MPNN's approach of focusing convolutions on bonds rather than atoms minimizes unnecessary computational loops during the message passing phase, thereby improving efficiency. The framework achieved strong results even on proprietary datasets, demonstrating the practical applicability and readiness of learned molecular representations in industrial settings.

Zhu et al. (2022) developed Hierarchical Informative Graph Neural Network (HiGNN) integrating molecular and fragment-level data. The model uses a decomposition algorithm and a feature-wise attention block for enhanced feature recalibration post-message passing. HiGNN's interpretability at the subgraph level, facilitated by a molecular-fragment similarity mechanism, helps with identifying crucial molecular components for design optimization.

Lusci et al. (2013) proposed building deep learning architectures using directed acyclic graphs (DAGs) to predict solubility with great results. This method involves utilizing a collection of RNNs linked to every possible vertex-centered acyclic orientation of a molecular graph. This reduces the need for feature engineering and allows the model to learn the best features given molecular descriptors.

2.4. Graph Attention Networks

GATs represent a significant advancement in the field of graph-based neural network models. Introduced by Veličković et al. (2018), GATs have addressed the limitations of earlier graph neural network architectures by incorporating attention mechanisms. The primary principle of GATs lies in their ability to assign varying levels of importance to nodes in a graph, thereby allowing for more nuanced feature aggregation.

GATs operate on the premise that not all nodes in a graph contribute equally to the representation of a given node. This is particularly pertinent in applications

where graph structures are irregular, making the uniform application of convolutional operations, as seen in traditional GCNs, less effective. By utilizing self-attention layers, GATs can weigh the influence of each node's neighbors, which leads to dynamic weight assignment based on the features of the nodes involved.

The development of GATs was a response to the need for more adaptive and flexible models that could handle complex, non-Euclidean data structures common in various real-world applications. One of the notable strengths of GATs is their applicability to a wide range of domains, including but not limited to social network analysis, recommendation systems, and biological data interpretation, particularly in understanding molecular structures and interactions.

In drug discovery contexts, GATs have shown promise in predicting drug interactions and drug efficacy by analyzing molecular structure graphs. The ability of GATs to focus on relevant parts of a graph makes them particularly suited for such tasks, where the significance of specific molecular substructures can vary greatly. The flexibility in capturing node dependencies without the need for complex matrix operations or extensive feature engineering underscores the practicality of GATs in handling intricate graph structures.

The output features of node i in GAT can be defined as:

$$h_i' = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W h_j\right)$$

In this expression, h_i' represents the updated feature vector of node i after

applying the attention mechanism. The sum iterates over the neighborhood $\mathcal{N}_i$ of node i, aggregating features (h_j) from neighbors weighted by the attention coefficients α_{ij}. These coefficients are learned during training and reflect the relative importance of each neighbor's features. W is a weight matrix applied to each neighbor's features before aggregation, and σ denotes a nonlinear activation function. This formula enables GATs to dynamically adjust to the information's relevance from different parts of the graph, enhancing their flexibility and power in capturing graph-embedded patterns.

2.5. Attentive FP

Attentive FP represent a novel approach in the field of cheminformatics, particularly in the domain of drug discovery and molecular property prediction. These networks leverage the principles of attention mechanisms, a concept borrowed from the realm of neural language processing, to enhance the ability of models to focus on relevant parts of molecular structures for predicting pharmacological properties.

The core idea behind Attentive FP is the utilization of a message-passing framework that enables the aggregation of information from both local atom-level features and global molecular-level features. This dual focus allows the network to learn not just the significance of individual atoms but also their contextual relevance within the entire molecular structure.

In Attentive FP, each atom has its own neighbor features that concentenate

both neighboring atoms and the connecting bond features Xiong et al. (2019). In each attentive layer, a novel state vector is produced for each atom. This state vector acquires neighborhood information as it moves through multiple successive attentive layers. The final state vector represents the learned structural information about the molecular graph, which is then followed by a layer designed for prediction.

2.6. Interpretability of GNNs

A common approach to explain predictions is by identifying subgraphs of the input graph with a small subset of node features that are most influential for the prediction. Ying et al. (2019) demonstrates this in their GNNEXPLAINER paper where they formulated as an optimization task that maximizes the mutual information between a GNN's prediction and distribution of possible subgraph structure

In the field of chemistry, (Wu et al., 2023) shows a very intuitive way of explaining graph neural networks for molecular property prediction with substructure tasking. They proposed substructure mask explanation (SME) based on established molecular segmentation methods. SMEs interpret GNN predictions based on molecular substructures, which makes more sense to chemists. By masking different substructures in a molecule to determine its influence in the model's ability to predict, this helps with explaining the model's prediction and makes it less of a black box.

There has been visual tools developed for GCN property prediction done by Kojima et al. (2020). Their tool, kGCN provides easy to use graphics user interface

for users to easily keep track and visualize the model training and evaluation process. This makes it easy to recognize the factors that affect the model's performance.

2.7. Ensemble Models in Machine Learning

Ensemble models in machine learning represent a paradigm where multiple learning algorithms are strategically employed to achieve better predictive performance than could be obtained from any of the constituent algorithms alone. This approach leverages the strength and mitigates the weaknesses of individual models, often leading to improved accuracy and robustness in predictions. In the context of chemical predictions, by combining diverse approaches such as decision trees, neural networks, or even multiple configurations of the same algorithm can provide a more nuanced understanding of these interactions.

The success of ensemble models depends on the diversity of the individual models, either by using different learning algorithms, different configurations of the same algorithm, or different subsets of training data. The models' individual predictions are then integrated using techniques such as voting, averaging, or stacking, where the outputs of individual models are input into a secondary model for final predictions.

Recent research highlights the evolving landscape of ensemble models in machine learning. Mohammed & Kora (2023) discuss various strategies in ensemble learning. Liu et al. (2021) introduced EGCN, which combines GCN with neural ar-

chitecture data processing to mitigate overfitting. Hu et al. (2021) applied ensemble models for predicting solubility, while Lu et al. (2019) developed a Multilevel Graph Convolutional Neural Network (MGCN) for molecular property prediction, emphasizing its generalizability and transferability. Lastly, Wang et al. (2023) experimented with a mixture of experts (MoE) approach in their GMoE model, demonstrating its efficacy in graph, node, and link prediction tasks.

2.8. Efficacy of GNNs in Various Contexts

Although GNNs have shown great promise in molecular activity prediction, they are not always the optimal choice. Descriptor-based models may be more suitable in some cases, especially with smaller datasets or simpler tasks. This notion is supported by a study conducted by Dejun et al. (2020), which analyzed the capability of eight machine learning (ML) algorithms, including four descriptor-based models (SVM, XGBoost, RF and DNN) and four graph-based models (GCN, GAT, MPNN and Attentive FP). They found that on average the descriptor based models performed better than the GNNs. However, in specific contexts, particularly with larger or multitask datasets, GNNs, such as GCN and Attentive FP could offer outstanding performance on a fraction of larger or multitask datasets.

In another study, Fung et al. (2021) also showed that descriptor-based models performing better than GNNs with small data sizes, but with GNNs performing better when ample data is available.

Chapter III.

Methodology

3.1. Computation

We used open source frameworks for machine learning, data handling, and optimization, including Ubuntu 22.04.3 LTS, DeepChem 2.7.2 (Ramsundar et al., 2019), Python 3.10.11, NumPy (Harris et al., 2020), Pandas (pandas development team, 2020), and pyGPGO (Jiménez-Luna & Ginebra, 2017). The experiments were conducted using a local NVIDIA RTX 3090 graphics card and Nvidia A10Gs on AWS. All experiments require no more than 24GB of video random-access memory (VRAM).

We implemented all models using the DeepChem library, using the `CSVLoader` class for data loading and the `MolGraphConvFeaturizer` for feature extraction. The models were instantiated with parameters optimized through Bayesian Optimization using the `pyGPGO` library, focused on maximizing the R^2 score while minimizing the MAE score.

Table 1 lists the key computing systems used in this work.

Type	Technology
Operating System	Ubuntu 22.04.3 LTS
Programming Language	Python 3.10.11
Machine Learning Framework	DeepChem
Numerical Computing Library	NumPy
Data Analysis Library	Pandas
Gaussian Process Optimization	pyGPGO
Logging	wandb
Hardware	GPU with 24GB VRAM

Table 1: Key Computing Systems and Libraries

3.2. Model Training, Validation, and Assessment

We trained GAT and Attentive FP networks on identical datasets, which were divided into training (80%), validation (10%), and testing (10%) sets. This approach aligns with previous research methodologies and adheres to the Pareto principle, optimizing the balance between training effectiveness and performance evaluation on unseen data.

Regarding hyperparameters, we tuned them using Bayesian Optimization, a method known for optimizing expensive black-box functions. We selected this approach due to its ability to handle noisy evaluations and its robustness in finding the global optimum. We then evaluate the models using our chosen metrics.

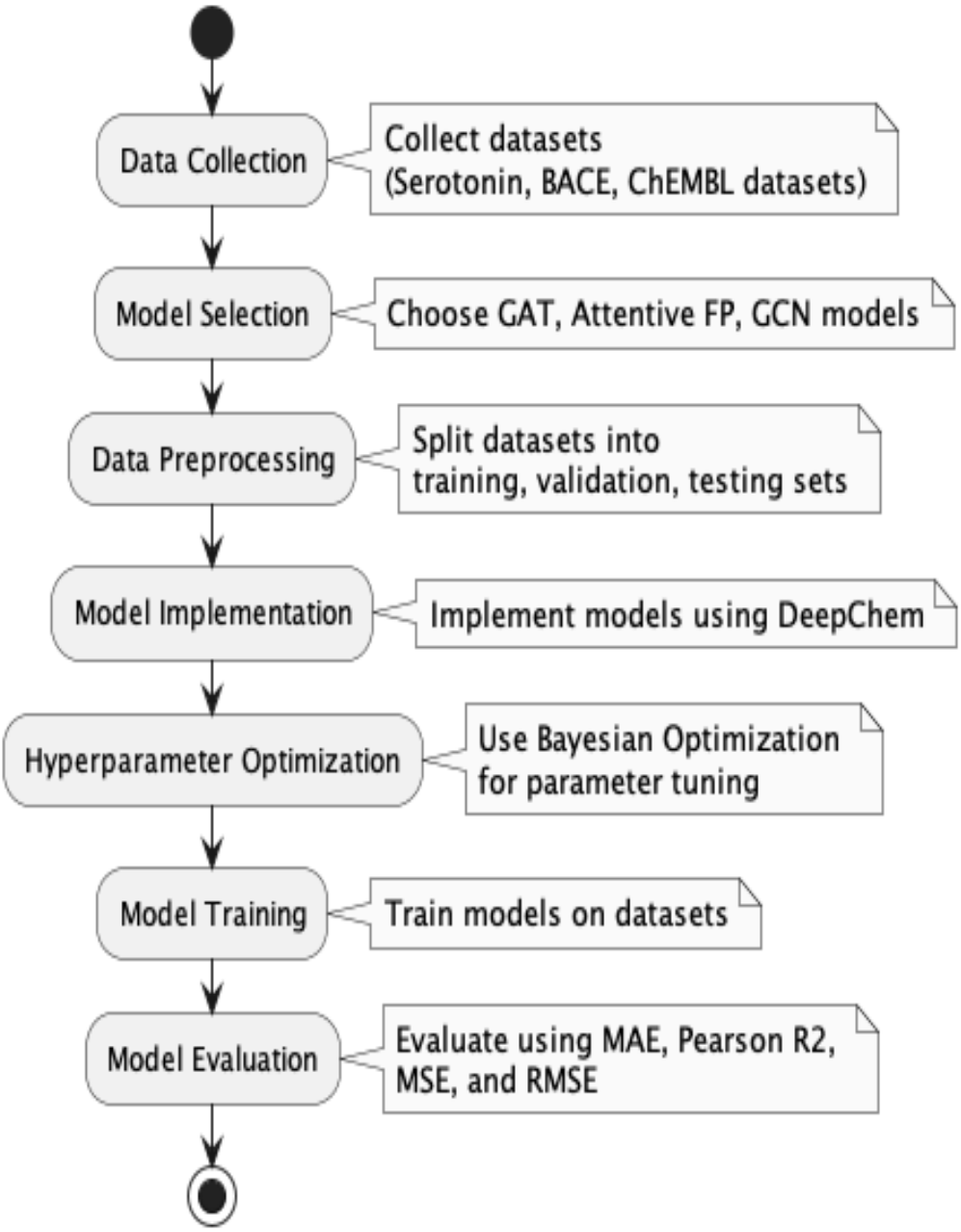

Figure 1: Flowchart of the Research Methodology

3.3. Data

We use four distinct datasets to investigate the performance of GAT and Attentive FP models. The datasets are described below:

(i) **Serotonin Transporter**: Provided by Sakai et al. (2021), this dataset contains 7890 compounds, each with a SMILES string and an affinity value.

(ii) **BACE**: From Wu et al. (2018), the BACE dataset comprises 1513 compounds and is used for regression analysis, focusing on beta-secretase 1 inhibitors relevant to Alzheimer's research.

(iii) **Acetylcholinesterase**: A curated selection of 422 compounds from ChEMBL, chosen for its smaller size to enable efficient model training and testing, focusing on acetylcholinesterase inhibitors.

(iv) **Target Protein Datasets**: Curated by Sakai et al. (2021), this dataset includes 127 targets from ChEMBL, offering a wide array of biological targets for benchmarking model performance.

3.4. Hyperparameter Optimization

We set custom ranges for our hyperparameters, including: graph attention layers, attention heads, dropout rate, learning rate, weight decay, and alpha. We evaluated them using the same $2R^2\_MAE$ metric defined by Sakai et al. (2021) in

Equation III.1. This selects for a higher R^2 score for parameter settings with the same MAE, leading to a better model fit.

$$2R^2\_MAE = (R^2 - \text{MAE}) + R^2 \tag{III.1}$$

3.4.1 GAT Hyperparameters

The hyperparameters for our GAT models, as detailed in Table 2, were established through a blend of empirical experimentation and literature analysis. These parameters include the size of graph attention layers, the number of attention heads, dropout rates, alpha values, predictor hidden features in the predictor, predictor dropout rates, learning rate, and weight decay.

Hyperparameter	**Type**	**Range**
Graph Attention Layer Size	int	[32, 2048]
Number of Attention Heads	int	[2, 16]
Dropout	continuous	[0.0, 0.5]
Alpha	continuous	[0.0, 1.0]
Predictor Hidden Features	int	[32, 512]
Predictor Dropout	continuous	[0.0, 0.5]
Learning Rate	continuous	[0.0001, 0.0020]
Weight Decay	continuous	[0.0, 0.01]

Table 2: Hyperparameters for Graph Attention Networks

3.4.2 Attentive FP Hyperparameters

For the Attentive FP models, we examined a set of hyperparameters detailed in Table 3. As with the GAT models, we customized ranges for the number of layers, timesteps, graph feature size, and dropout rates.

Hyperparameter	**Type**	**Range**
Learning Rate	continuous	[0.0001, 0.0020]
Number of Layers	integer	[1, 5]
Number of Timesteps	integer	[1, 5]
Graph Feature Size	integer	[100, 300]
Dropout	continuous	[0.0, 0.5]

Table 3: Hyperparameters for Attentive FP Networks

Chapter IV.

Results

4.1. Evaluation of Models Across Datasets

4.1.1 Serotonin Transporter

The Serotonin Transporter Dataset contains the SMILE strings of 7890 compounds. Sakai et al. (2021) used this dataset for virtual screening and identified a new compound with activity similar to that of a marketed drug in in vivo assays. We selected this dataset as our initial point of comparison between GATs and Attentive FP against Sakai's custom GCN model. We performed the hyperparameter optimization process mentioned in 3.4 and applied identical settings to all three models. The results are summarized in Table 4.

Table 4: Consolidated Performance Metrics on Serotonin Transporter

Metric	Method	Training	Validation	Testing
MAE	GCN	0.439375	0.553131	0.568916
MSE	GCN	0.571541	0.724126	0.739575
R^2	GCN	0.823894	0.721155	0.701071
MAE	GAT	0.261709	0.625375	0.602776
MSE	GAT	0.352881	0.820818	0.798426
R^2	GAT	0.932803	0.631962	0.668463
MAE	Attentive FP	0.258687	0.544640	0.551442
MSE	Attentive FP	0.351476	0.726672	0.729712
R^2	Attentive FP	0.940689	0.727299	0.723846

We observed that Attentive FP surpassed other models in terms of MAE and R^2 metrics. It is worth noting that both Attentive FP and GAT exhibited a significantly larger performance gap between the training and validation sets compared to GCN, suggesting that the graph attention-based networks may have overtrained on nonlocal properties. Such overfitting indicates a potential limitation in their ability to model simpler interactions across diverse datasets. However, it could be easily mitigated by incorporating regularization techniques or augmenting the training data to cover a broader range of scenarios.

4.1.2 BACE

We proceeded to test our models on the BACE dataset (Wu et al., 2018), The outcomes of these evaluations are documented in Table 5.

Table 5: Consolidated Performance Metrics on BACE

Metric	Method	Training	Validation	Testing
MAE	GCN	0.188240	0.400432	0.430842
MSE	GCN	0.251903	0.574059	0.541392
R^2	GCN	0.942777	0.716605	0.703539
MAE	GAT	0.360441	0.450633	0.434854
MSE	GAT	0.459834	0.597393	0.554088
R^2	GAT	0.866507	0.727799	0.783553
MAE	Attentive FP	0.094698	0.432758	0.337523
MSE	Attentive FP	0.163809	0.585793	0.469685
R^2	Attentive FP	0.975934	0.703627	0.797840

Attentive FP demonstrated superior performance across all metrics, surpassing both GAT and GCN. While Attentive FP and GCNs exhibited strong results on the training set, their effectiveness was less notable on the validation and testing sets, which might be due to the small size of the dataset. In this instance, GATs showed a higher R^2 value on the testing set yet remained inferior to GCNs in terms of MAE. Nonetheless, the marginal difference between GAT and GCN suggests that the models

may be used interchangeably.

4.1.3 Acetylcholinesterase

The Acetylcholinesterase dataset was derived by further filtering the ChEMBL 24 dataset, originally curated by Bosc et al. (2019). We found that most targets from the paper had between 50 and 200 datapoints. We opted for the largest subset, the Acetylcholinesterase target with 422 data points. Similar to previous experiments, Attentive FP demonstrated superior performance over the other two models in R^2 and MAE metrics. Although GATs showed slightly higher R^2 scores, their performance was lower in both MAE and MSE metrics. These results are summarized in Table 6.

Table 6: Consolidated Performance Metrics on Acetylcholinesterase

Metric	Method	Training	Validation	Testing
MAE	GCN	0.331082	0.438151	0.529737
MSE	GCN	0.446471	0.588697	0.684809
R^2	GCN	0.830635	0.748961	0.596156
MAE	GATs	0.250314	0.545306	0.577207
MSE	GATs	0.327429	0.698976	0.760359
R^2	GATs	0.925153	0.590297	0.629249
MAE	Attentive FP	0.055869	0.481681	0.477356
MSE	Attentive FP	0.110144	0.703513	0.618998
R^2	Attentive FP	0.989614	0.618218	0.752217

Interestingly, we observed that the R^2 score showed a noticeable increase from the validation to the testing sets for GAT, and even more significantly for Attentive FP. This outcome is unusual, though not unprecedented, as typically scores deteriorate when transitioning from validation to testing. We hypothesize that this anomaly may be attributed to random variation and possibly the quality and size of the dataset. Given that this is already the largest subset, we opted not to extend the experiment to even smaller subsets. Nonetheless, according to Sakai et al. (2021), a model is considered a good model if it achieves either an MAE < 0.6 or an $R^2 > 0.6$. In this case, all models met these criteria.

4.1.4 Target Protein Datasets

To broaden the evaluation of GAT and Attentive FP, we leveraged the datasets encompassing 127 target proteins as provided in Sakai et al. (2021). We trained our models and evaluated their performance on these datasets. In Table 7, we present the targets corresponding to the top 4 performing models from Sakai's study, selected based on the lowest MAE scores. The full results are summarized in Table 8 and Table 9.

Protein Family	Target	Size	GCN		Attentive FP		GAT	
			MAE	R^2	MAE	R^2	MAE	R^2
GPCR	Orexin Receptor 1	2852	0.41 ± 0.013	0.79	0.497706	0.591447	0.475512	0.636221
GPCR	Serotonin 7 5-HT7 Receptor	2395	0.47 ± 0.023	0.74	0.626231	0.565231	0.62795	0.528161
GPCR	Orexin Receptor 2	3079	0.50 ± 0.010	0.71	0.597953	0.518812	0.526359	0.612372
GPCR	Cannabinoid CB1 Receptor	6966	0.51 ± 0.0080	0.76	0.588262	0.604333	0.584363	0.615557
Enzyme	Acetyl-CoA Carboxylase 2	3136	0.33 ± 0.018	0.68	0.339897	0.59946	0.348305	0.59923
Enzyme	Poly ADP-Ribose Polymerase-1	3101	0.42 ± 0.012	0.82	0.477657	0.72486	0.475141	0.737627
Enzyme	Cholinesterase	3011	0.43 ± 0.015	0.82	0.464326	0.755615	0.442243	0.759405
Enzyme	Nicotinamide Phosphoribosyltransferase	2342	0.45 ± 0.011	0.68	0.460304	0.552548	0.431985	0.623034
Ion Channel	hERG	9198	0.42 ± 0.013	0.66	0.402358	0.585391	0.477247	0.524965
Ion Channel	Voltage-Gated Potassium Channel Subunit Kv1.5	739	0.42 ± 0.020	0.53	0.488535	0.430439	0.497179	0.458191
Ion Channel	Sodium Channel Protein Type IX Alpha Subunit	5677	0.47 ± 0.016	0.72	0.555087	0.577707	0.481973	0.648225
Ion Channel	Vanilloid Receptor	2856	0.50 ± 0.017	0.78	0.5731	0.609582	0.541078	0.684196
Kinase	Nerve Growth Factor Receptor Trk-A	2587	0.42 ± 0.017	0.71	0.44961	0.650777	0.471246	0.616584
Kinase	Insulin-Like Growth Factor I Receptor	3019	0.44 ± 0.010	0.85	0.457185	0.811463	0.509064	0.784112
Kinase	Tyrosine-Protein Kinase JAK1	4345	0.45 ± 0.012	0.81	0.514207	0.657184	0.550293	0.636757
Kinase	Serine Threonine-Protein Kinase mTOR	4414	0.46 ± 0.018	0.81	0.48659	0.751609	0.478746	0.752173
Nuclear Receptor	Thyroid Hormone Receptor Alpha	461	0.40 ± 0.014	0.82	0.317999	0.905391	0.432221	0.838262
Nuclear Receptor	Glucocorticoid Receptor	2293	0.53 ± 0.026	0.78	0.60594	0.647011	0.606133	0.650094
Nuclear Receptor	Peroxisome Proliferator-Activated Receptor Gamma	3018	0.55 ± 0.015	0.72	0.587907	0.665589	0.573776	0.680704
Nuclear Receptor	Vitamin D Receptor	546	0.54 ± 0.030	0.88	0.501735	0.853711	0.552683	0.819698
Protease	Cathepsin D	2568	0.42 ± 0.018	0.85	0.501099	0.821792	0.501263	0.804864
Protease	Matrix Metalloproteinase-1	3746	0.47 ± 0.020	0.81	0.455911	0.63966	0.468225	0.652845
Protease	ADAM17	2410	0.47 ± 0.022	0.89	0.457197	0.861113	0.510525	0.849666
Protease	Cathepsin S	2309	0.50 ± 0.010	0.79	0.597208	0.681064	0.604124	0.689075
Transporter	Potassium-Transporting ATPase	532	0.42 ± 0.0081	0.52	0.518178	0.546349	0.49603	0.587555
Transporter	GABA Transporter 1	576	0.47 ± 0.040	0.86	0.477037	0.783785	0.495396	0.743078
Transporter	Dopamine Transporter	5908	0.54 ± 0.014	0.76	0.670739	0.656383	0.55616	0.737176
Transporter	Norepinephrine Transporter	4342	0.55 ± 0.015	0.7	0.631822	0.544152	0.618925	0.53601
Others	Histone Deacetylase 1	4239	0.47 ± 0.015	0.74	0.48994	0.659453	0.497822	0.670272
Others	Bromodomain-Containing Protein 4	2208	0.46 ± 0.032	0.82	0.53122	0.694395	0.528409	0.710369
Others	Histone Deacetylase 6	2725	0.47 ± 0.023	0.82	0.564483	0.711331	0.522767	0.734411
Others	P53-Binding Protein MDM-2	2346	0.47 ± 0.020	0.88	0.48738	0.866853	0.466591	0.883679

Table 7: Comparing GAT and Attentive FP against Sakai's Top 4 GCN Models for Each Protein Family

Our results indicate that many of our GAT and Attentive FP models achieved MAE scores that fall within the range of GCNs reported by Sakai et al. (2021). Also, 30 out of 32 models from both GAT and Attentive FP categories met Sakai's criteria of having MAE < 0.6 or an $R^2 > 0.6$. Considering all 127 models, 104 GAT models and 98 Attentive FP models met this criterion.

GATs also achieved a better R^2 fit in P53 Binding Protein MDM-2, Potassium Transporting ATPase, and Thyroid Hormone Receptor Alpha. Similarly, Attentive FP showed better R^2 values in Potassium Transporting ATPase and Thyroid Hormone Receptor Alpha. Particular striking was the superb performance of Attentive FP models on the Thyroid Hormone Receptor Alpha dataset, which, containing only 461 data points, is the smallest in our study. This finding prompted us to further investigate the relationship between dataset size and model performance, a topic we explore in Section 4.3.

4.2. Our Top 4 Models for Each Protein Family

We further evaluated our models on the complete set of 127 target protein datasets provided by Sakai, identifying the top four models for each protein family based on MAE scores. The summarized results are presented in Table 8.

Protein Family	Target	MAE	R^2
GPCR	Corticotropin Releasing Factor Receptor 1	0.490922	0.623825
GPCR	Dopamine D3 Receptor	0.531025	0.663432
GPCR	Orexin Receptor 1	0.53926	0.553303
GPCR	Serotonin 6 5-HT6 Receptor	0.556455	0.612263
Enzyme	Acetyl-CoA Carboxylase 2	0.339142	0.604616
Enzyme	Nicotinamide Phosphoribosyltransferase	0.424282	0.6268
Enzyme	Poly ADP-Ribose Polymerase-1	0.473701	0.735631
Enzyme	Arachidonate 5-Lipoxygenase	0.506763	0.53998
Ion Channel	hERG	0.453352	0.519903
Ion Channel	Voltage-Gated Potassium Channel Subunit Kv1.5	0.510276	0.423206
Ion Channel	P2X Purinoceptor 7	0.586491	0.360953
Ion Channel	Vanilloid Receptor	0.618081	0.598255
Kinase	Serine Threonine-Protein Kinase mTOR	0.43832	0.780209
Kinase	Fibroblast Growth Factor Receptor 3	0.447547	0.681579
Kinase	Fibroblast Growth Factor Receptor 1	0.458142	0.823691
Kinase	Insulin-Like Growth Factor I Receptor	0.468389	0.790613
Nuclear Receptor	Thyroid Hormone Receptor Alpha	0.403229	0.80651
Nuclear Receptor	Vitamin D Receptor	0.538696	0.848425
Nuclear Receptor	Androgen Receptor	0.572326	0.676986
Nuclear Receptor	Peroxisome Proliferator-Activated Receptor Gamma	0.602087	0.642367
Protease	Cathepsin D	0.453734	0.855205
Protease	Matrix Metalloproteinase-1	0.472195	0.639366
Protease	ADAM17	0.497856	0.82213
Protease	Leukocyte Elastase	0.567666	0.946549
Transporter	GABA Transporter 1	0.45197	0.786631
Transporter	Serotonin Transporter	0.541849	0.710211
Transporter	Norepinephrine Transporter	0.574207	0.602269
Transporter	Dopamine Transporter	0.594363	0.717012
Others	Histone Deacetylase 1	0.479778	0.68022
Others	P53-Binding Protein MDM-2	0.488922	0.879627
Others	Bromodomain-Containing Protein 4	0.504837	0.751841
Others	Apoptosis Regulator BCL-2	0.568045	0.829819

Table 8: Top 4 Attentive FP Models for Each Protein Family

The targets that appeared in our top 4 Attentive FP models did not always match those in Sakai's top 4 models. The Protease family showed the most overlap, with the top 3 models appearing in Sakai's rankings in the same order. For other families, only 1 or 2 targets were common across both rankings. This variation suggests that model performance may depend on the target protein, indicating that the best model for one target might not be the best for another. A similar analysis was conducted for GATs, with the results summarized in Table 9. In the case of GATs, while the Protease family again showed the most overlap with 3 common targets, their order differed from Sakai's rankings.

Protein Family	Target	MAE	R^2
GPCR	Orexin Receptor 1	0.475512	0.636221
GPCR	Corticotropin Releasing Factor Receptor 1	0.515567	0.65453
GPCR	Orexin Receptor 2	0.526359	0.612372
GPCR	Adenosine A2A Receptor	0.551854	0.678465
Enzyme	Acetyl-CoA Carboxylase 2	0.348305	0.59923
Enzyme	Nicotinamide Phosphoribosyltransferase	0.431985	0.623034
Enzyme	Poly ADP-Ribose Polymerase-1	0.475141	0.737627
Enzyme	Protein-Tyrosine Phosphatase 1B	0.494986	0.656528
Ion Channel	hERG	0.477247	0.524965
Ion Channel	Sodium Channel Protein Type IX Alpha Subunit	0.481973	0.648225
Ion Channel	Voltage-Gated Potassium Channel Subunit Kv1.5	0.497179	0.458191
Ion Channel	Transient Receptor Potential Cation Channel Subfamily M Member 8	0.519204	0.828398
Kinase	Fibroblast Growth Factor Receptor 3	0.447973	0.706114
Kinase	Nerve Growth Factor Receptor Trk-A	0.471246	0.616584
Kinase	Serine Threonine-Protein Kinase mTOR	0.478746	0.752173
Kinase	Serine Threonine-Protein Kinase B-Raf	0.488564	0.734272
Nuclear Receptor	Thyroid Hormone Receptor Alpha	0.432221	0.838262
Nuclear Receptor	Androgen Receptor	0.509433	0.741759
Nuclear Receptor	Vitamin D Receptor	0.552683	0.819698
Nuclear Receptor	Peroxisome Proliferator-Activated Receptor Gamma	0.573776	0.680704
Protease	Matrix Metalloproteinase-1	0.468225	0.652845
Protease	Cathepsin D	0.501263	0.804864
Protease	ADAM17	0.510525	0.849666
Protease	Beta-Secretase 1	0.590782	0.635515
Transporter	GABA Transporter 1	0.495396	0.743078
Transporter	Potassium-Transporting ATPase	0.49603	0.587555
Transporter	Dopamine Transporter	0.55616	0.737176
Transporter	Serotonin Transporter	0.566232	0.685962
Others	P53-Binding Protein MDM-2	0.466591	0.883679
Others	Histone Deacetylase 1	0.497822	0.670272
Others	Histone Deacetylase 6	0.522767	0.734411
Others	Bromodomain-Containing Protein 4	0.528409	0.710369

Table 9: Top 4 GAT Models for Each Protein Family

4.2.1 Comparison of GAT and Attentive FP

We display the per-family MAE and R^2 scores for all 127 targets in Figure 2. The data reveal that the families showing the most significant differences in performance between the two models are in the GPCR, Kinase, and Nuclear Receptor families. Specifically, GATs significantly outperform Attentive FP within the GPCR and Kinase families, whereas Attentive FP demonstrates superior performance in the Nuclear Receptor family.

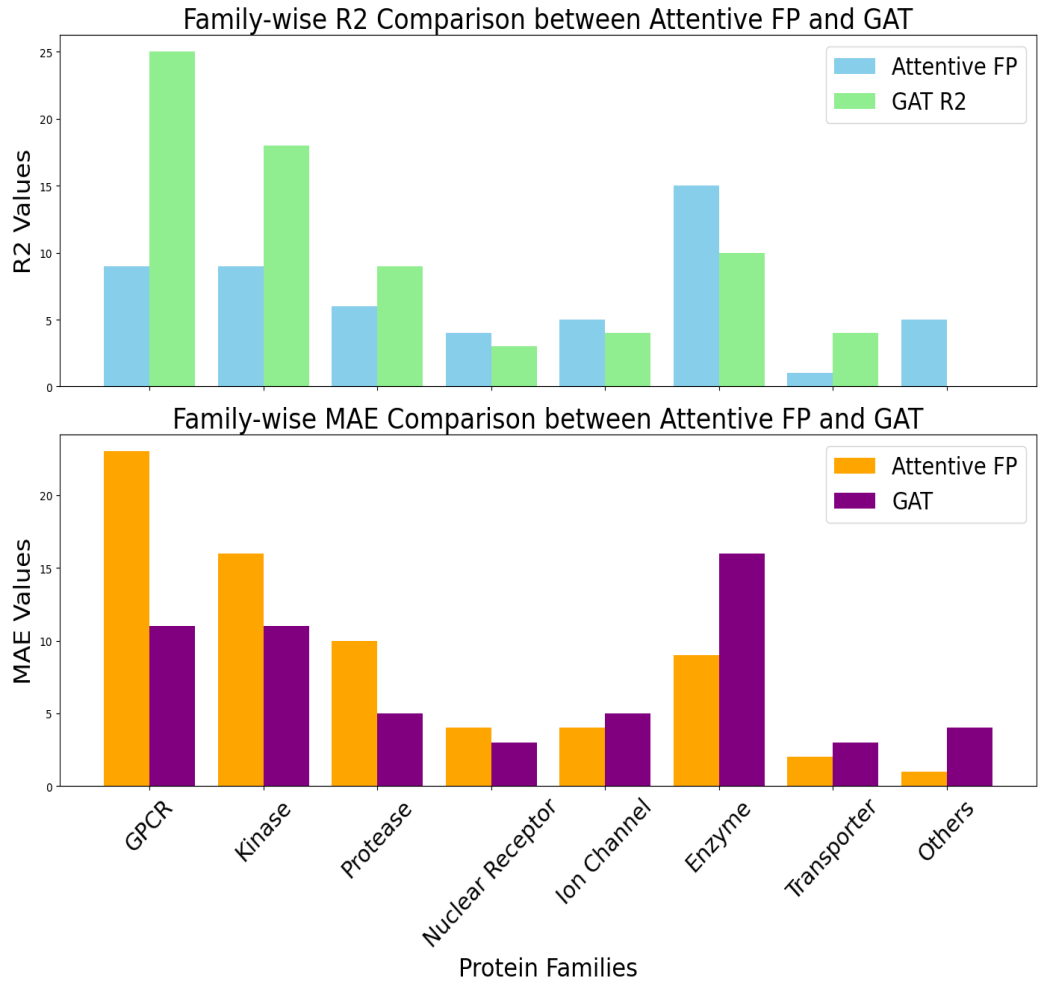

Figure 2: Performance of GAT and Attentive FP on 127 Target Protein Datasets

4.3. Experiments on Truncated Datasets

In the evaluated datasets, GATs and Attentive FP demonstrated good performance on the Serotonin Transporter, BACE, and Acetylcholinesterase datasets compared to GCNs. According to the standards established by Sakai et al. (2021), most of our models qualified as good models. We observed that model performance not only varied across datasets with different target families but also dataset sizes. We speculate that some of this performance variability may be linked to the differences in the sizes of the datasets. For example, the Serotonin Transporter dataset includes 7890 data points, whereas the BACE and Acetylcholinesterase datasets contain 1513 and 422 data points. Our GAT and Attentive FP models outperformed Sakai's GCN model on the Thyroid Hormone Receptor Alpha dataset, which has only 461 data points. To delve deeper, we truncate all target protein dataset sizes to 461 data points, the size of the smallest dataset, and reassessed our models. The results of this experiment are presented in Section 4.3.

4.3.1 Attentive FP Model Performance

We discovered that for the Protease, Transporter, and Enzyme families, three of the top four models remained consistent compared to their performance before truncation. This consistency suggests that these families might be more resilient to variations in dataset size. Notably, ADAM17 showed minimal performance decline after its reduction from 2410 to 461 data points. For Transporters and Nuclear

Receptors, half of the top four models still qualify as effective, although a significant performance reduction was observed in most targets. These findings are detailed in Table 10.

Protein Family	Target	MAE	R^2
GPCR	Dopamine D1 Receptor	0.682311	0.452283
GPCR	Orexin Receptor 2	0.700794	0.290899
GPCR	G Protein-Coupled Receptor 44	0.701444	0.388748
GPCR	Sigma Opioid Receptor	0.728259	0.298365
Enzyme	Acetyl-CoA Carboxylase 2	0.517972	0.231402
Enzyme	Nicotinamide Phosphoribosyltransferase	0.647807	0.172251
Enzyme	Cyclooxygenase-1	0.663365	0.198229
Enzyme	Poly ADP-Ribose Polymerase-1	0.720633	0.547431
Ion Channel	Voltage-Gated Potassium Channel Subunit Kv1.5	0.608806	0.371786
Ion Channel	Transient Receptor Potential Cation Channel Subfamily M Member 8	0.614203	0.784754
Ion Channel	Sodium Channel Protein Type IX Alpha Subunit	0.694951	0.329833
Ion Channel	Transient Receptor Potential Cation Channel Subfamily A Member 1	0.708502	0.396932
Kinase	Serine Threonine-Protein Kinase Aurora-A	0.602185	0.744449
Kinase	Fibroblast Growth Factor Receptor 3	0.622104	0.544021
Kinase	Nerve Growth Factor Receptor Trk-A	0.63105	0.294513
Kinase	Tyrosine-Protein Kinase Receptor Flt3	0.686655	0.478522
Nuclear Receptor	Thyroid Hormone Receptor Alpha	0.353245	0.875037
Nuclear Receptor	Vitamin D Receptor	0.718436	0.735657
Nuclear Receptor	Glucocorticoid Receptor	0.780833	0.423534
Nuclear Receptor	Estrogen Receptor Alpha	0.815322	0.467774
Protease	ADAM17	0.5552	0.812146
Protease	Matrix Metalloproteinase-1	0.69898	0.463085
Protease	Cathepsin D	0.722064	0.539881
Protease	Matrix Metalloproteinase 9	0.727469	0.654812
Transporter	GABA Transporter 1	0.532831	0.68282
Transporter	Potassium-Transporting ATPase	0.551369	0.555607
Transporter	Norepinephrine Transporter	0.809101	0.369478
Transporter	Dopamine Transporter	0.929396	0.356622
Others	Apoptosis Regulator BCL-2	0.640531	0.755704
Others	Bromodomain-Containing Protein 4	0.686667	0.461269
Others	Histone Deacetylase 1	0.739618	0.329351
Others	Histone Deacetylase 6	0.860364	0.345978

Table 10: Attentive FP Performance Metrics on truncated Target Protein Datasets with 461 Datapoints (Top 4)

4.3.2 GAT Performance

We also evaluated GATs on the truncated datasets. Our findings indicate that for the Ion Channel and Transporter families, three out of the top four targets remained the same, whereas for the Protease family, all four targets persisted, though with a different order. ADAM17 maintained its strong performance following the truncation. Moreover, 13 GATs achieved a MAE of less than 0.6 or R^2 greater than 0.6, compared to 10 models for Attentive FP. These results suggest a greater resilience of GATs to variations in dataset size relative to Attentive FP. The detailed outcomes are presented in Table 11.

Protein Family	Target	MAE	R^2
GPCR	Orexin Receptor 2	0.589704	0.452898
GPCR	Dopamine D1 Receptor	0.60931	0.486423
GPCR	Orexin Receptor 1	0.612161	0.580473
GPCR	Cholecystokinin B Receptor	0.668391	0.694264
Enzyme	Acetyl-CoA Carboxylase 2	0.424239	0.233817
Enzyme	Nicotinamide Phosphoribosyltransferase	0.599874	0.208696
Enzyme	Carbonic Anhydrase XII	0.669231	0.422563
Enzyme	Poly ADP-Ribose Polymerase-1	0.742217	0.566515
Ion Channel	Sodium Channel Protein Type IX Alpha Subunit	0.642392	0.402947
Ion Channel	Transient Receptor Potential Cation Channel Subfamily A Member 1	0.642817	0.548105
Ion Channel	Voltage-Gated Potassium Channel Subunit Kv1.5	0.6506	0.401565
Ion Channel	P2X Purinoceptor 7	0.669577	0.218545
Kinase	Tyrosine-Protein Kinase SYK	0.550226	0.501422
Kinase	Nerve Growth Factor Receptor Trk-A	0.562369	0.319242
Kinase	Serine Threonine-Protein Kinase B-Raf	0.679553	0.546901
Kinase	PI3-Kinase P110-Delta Subunit	0.698697	0.49173
Nuclear Receptor	Thyroid Hormone Receptor Alpha	0.432221	0.838262
Nuclear Receptor	Vitamin D Receptor	0.676815	0.747461
Nuclear Receptor	Glucocorticoid Receptor	0.70666	0.562693
Nuclear Receptor	Estrogen Receptor Alpha	0.774727	0.600589
Protease	Cathepsin D	0.57719	0.67336
Protease	ADAM17	0.684251	0.803204
Protease	Matrix Metalloproteinase-1	0.720367	0.490284
Protease	Beta-Secretase 1	0.733293	0.357651
Transporter	Potassium-Transporting ATPase	0.489583	0.553208
Transporter	GABA Transporter 1	0.539342	0.692818
Transporter	Norepinephrine Transporter	0.827091	0.4493
Transporter	Dopamine Transporter	0.97181	0.268033
Others	Bromodomain-Containing Protein 4	0.605324	0.568015
Others	Apoptosis Regulator BCL-2	0.704223	0.75977
Others	Histone Deacetylase 6	0.727624	0.555039
Others	Histone Deacetylase 1	0.878685	0.233667

Table 11: GAT Performance Metrics on truncated Target Protein Datasets with 461 Datapoints (Top 4)

4.3.3 Comparison of GAT and Attentive FP

Attentive FP no longer falls behind in the number of models with superior R^2 scores for GPCR and Kinases. It also retains its lead in Nuclear Receptors, as illustrated in Figure 3. These results suggest that Attentive FP might exhibit greater resilience to dataset size variations compared to GAT, particularly in smaller datasets. This observation is consistent with our previous findings, where Attentive FP achieved significantly higher scores in its training and validation sets than in testing.

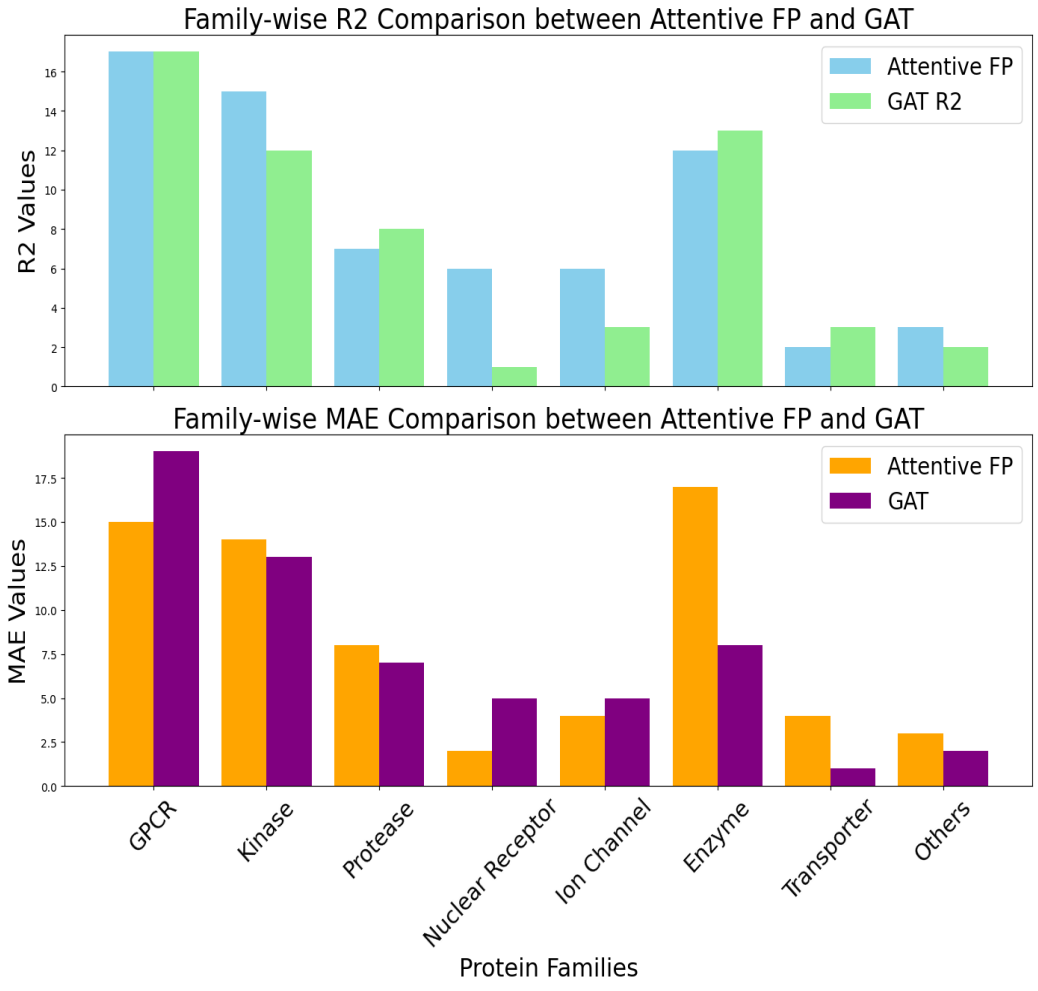

Figure 3: Comparison of GAT and Attentive FP on Truncated Datasets

4.3.4 GCN Performance on Truncated Datasets

We also trained Sakai's GCN model on the truncated dataset, with the results detailed in Table 12. In alignment with the outcomes from GAT and Attentive FP evaluations, the Protease and Transporter families demonstrated considerable resilience to the reduction in dataset size, maintaining consistency in three of the top four models. Additionally, 14 GCN models achieved an MAE < 0.6 or an $R^2 > 0.6$, aligning with our earlier observations that GCNs may be less prone to overfitting relative to GAT and Attentive FP. We also observed that GABA transporter 1 showed significant improvements in both MAE and R^2, suggesting that the reduction in dataset size may have eliminated extraneous noise, thereby enhancing the model's predictive accuracy.

Protein Family	Target	MAE	R^2
GPCR	Corticotropin Releasing Factor Receptor 1	0.53043	0.613273
GPCR	Orexin Receptor 1	0.563273	0.501651
GPCR	Orexin Receptor 2	0.648319	0.58323
GPCR	Dopamine D4 Receptor	0.68916	0.298028
Enzyme	Acetyl-CoA Carboxylase 2	0.531899	0.190923
Enzyme	Cyclooxygenase-1	0.643148	0.005915
Enzyme	Nicotinamide Phosphoribosyltransferase	0.655247	0.452706
Enzyme	Carbonic Anhydrase XII	0.679434	0.514694
Ion Channel	Voltage-Gated Potassium Channel Subunit Kv1.5	0.521221	0.576173
Ion Channel	Transient Receptor Potential Cation Channel Subfamily A Member 1	0.637406	0.563163
Ion Channel	Transient Receptor Potential Cation Channel Subfamily M Member 8	0.693393	0.596989
Ion Channel	P2X Purinoceptor 7	0.731497	0.215338
Kinase	Fibroblast Growth Factor Receptor 1	0.591589	0.668814
Kinase	Fibroblast Growth Factor Receptor 3	0.603639	0.2777
Kinase	PI3-Kinase P110-Gamma Subunit	0.61724	0.453256
Kinase	Tyrosine-Protein Kinase SYK	0.623505	0.657596
Nuclear Receptor	Thyroid Hormone Receptor Alpha	0.367244	0.832336
Nuclear Receptor	Vitamin D Receptor	0.507467	0.868099
Nuclear Receptor	Peroxisome Proliferator-Activated Receptor Gamma	0.689822	0.54366
Nuclear Receptor	Estrogen Receptor Alpha	0.69876	0.657512
Protease	Cathepsin D	0.739058	0.608501
Protease	Cathepsin S	0.739394	0.600441
Protease	Matrix Metalloproteinase-1	0.770415	0.367968
Protease	Beta-Secretase 1	0.781328	0.427122
Transporter	GABA Transporter 1	0.296487	0.934026
Transporter	Potassium-Transporting ATPase	0.62387	0.389884
Transporter	Norepinephrine Transporter	0.687844	0.48581
Transporter	Dopamine Transporter	0.881828	0.273499
Others	P53-Binding Protein MDM-2	0.628305	0.73832
Others	Bromodomain-Containing Protein 4	0.660821	0.703937
Others	Apoptosis Regulator BCL-2	0.750484	0.768536
Others	Histone Deacetylase 1	0.776877	0.249638

Table 12: GCN Performance Metrics on Truncated Target Protein Datasets with 461 Datapoints (Top 4)

Chapter V.

Discussions

5.1. Performance of GNNs across different protein families

Our analysis shows that Attentive FP consistently outperformed GATs and GCNs across the Serotonin Transporter, BACE, and Acetylcholinesterase datasets. While GATs were also competitive on the BACE dataset, their performance was mixed on Sakai's 127 target protein dataset.

When comparing the top four targets by MAE against Sakai's best models, Attentive FP had three overlaps in the rankings for Enzymes, Ion Channels, Proteases, Transporters, and "Others" categories, two for Kinases and Nuclear Receptors, and one for GPCRs. For GATs, all four matched in "Others," three in Enzymes, Ion Channels, Proteases, and Transporters, two in GPCRs and Kinases, and one in Nuclear Receptors. Despite the overlaps, there are notable ranking differences between GAT and Attentive FP when applied to the same datasets. For instance, in Kinases, while only two top targets overlapped, Attentive FP's top four MAEs were comparable to those of GCNs, as illustrated in Figure 4.

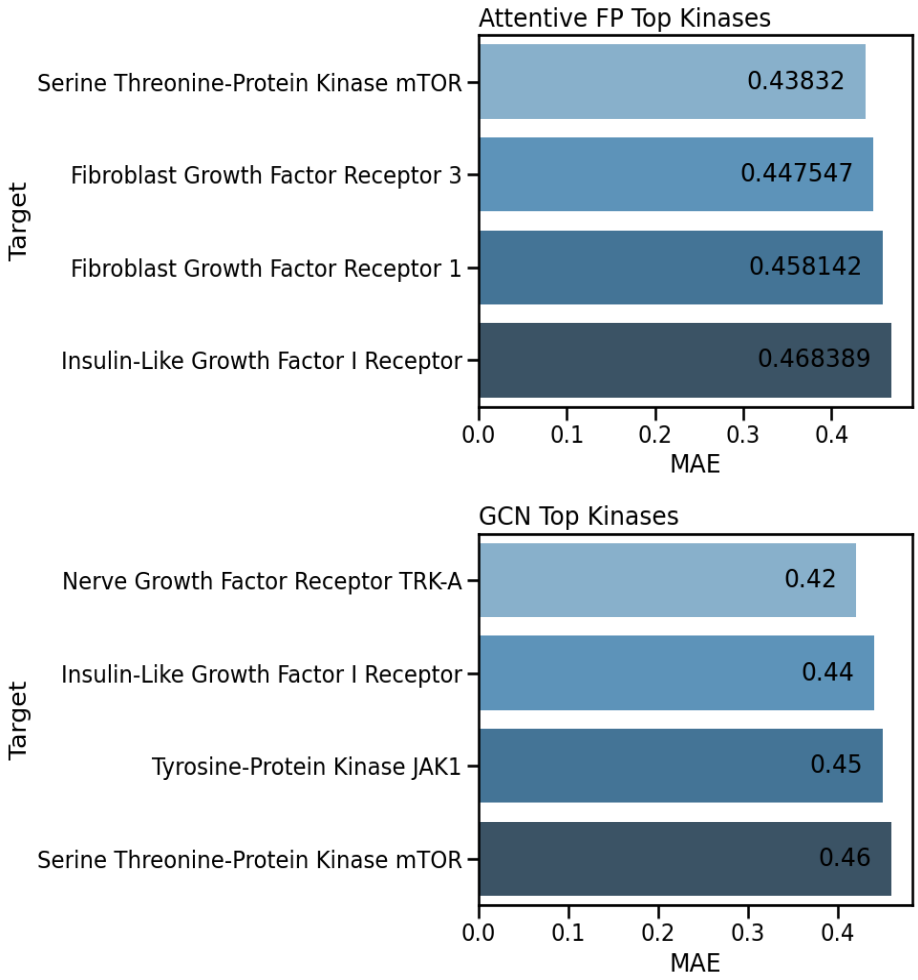

Figure 4: Top 4 models for Kinases using Attentive FP and GCNs

The findings suggest that Attentive FP could be a viable alternative for binding affinity prediction tasks in Kinases, similar to GATs in many other protein families. This variability underscores the importance of model selection based on the dataset and the specific target protein family, echoing the conclusions of McCardle (2023), which noted minimal differences in predictive capabilities among GNN architectures across diverse datasets. Our study reinforces the notion that while model diversity may have less impact on predictive accuracy than previously thought, the character-

istics of target protein families are crucial in determining the most effective modeling approach.

5.2. GATs and Attentive FP

There is a noticeable performance gap between Attentive FP and GATs, particularly in the GPCR and Kinase families, as detailed in Figure 5. GATs outperform Attentive FP in these categories, showing higher counts of better R^2 scores. However, there are individual instances where Attentive FP significantly exceeds GATs performance in these families. In other protein families, the performance between the two models is comparable, with Attentive FP often having a slight advantage.

The observed performance gap between Attentive FP and GATs in those families may be attributed to the different ways these models handle graph structures and node features. Attentive FP allows for more nuanced weighting of interactions, which could be particularly beneficial for the complex binding sites found in these protein families. However, this also contributes to the increased computational time, suggesting a trade-off between predictive performance and efficiency that must be considered in practical applications.

Our preliminary tests indicate that Attentive FP could take up to 25% longer to run. Despite this, while Attentive FP can excel under certain conditions, GATs generally serve as a more reliable first choice for many datasets.

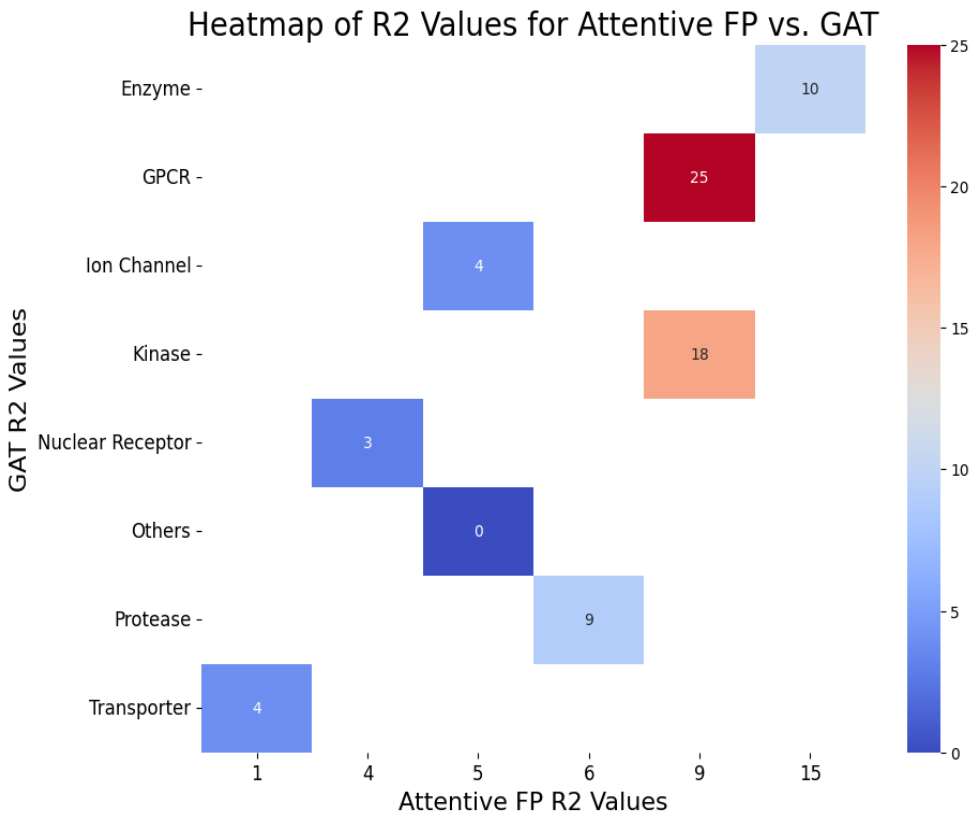

Figure 5: Heatmap of the performance of GAT and Attentive FP across different datasets

5.3. Impact of truncating datasets

Overall, the truncation of datasets negatively impacted the performance of the models, as the number of good models using Sakai's heuristic of $R^2 > 0.6$ or $MAE < 0.6$ decreased from 104 to 13 for GAT and 98 to 10 for Attentive FP. However, while most of the models generally performed worse when the datasets were truncated, the few that remained the same or improved shows that for certain targets, the models were able to generalize well.

For example, we found that Proteases and Transporters generally maintained

the same models in the top 4 when ranked by MAE as shown earlier in Figure 6 and 7. In the following sections, we try to take a deeper look into the impact of truncation for GAT and Attentive FP by taking a look at the families that did change, the magnitude of changes in our metrics, and the few targets that saw an increase in performance.

5.3.1 GAT

GAT models generally performed worse after truncation. The Enzyme family was very sensitive to the change in dataset size, while Nuclear Receptors were the most resilient. This mostly aligns with the results from Attentive FP. Overall, it appears that dataset truncation has had a larger impact on the performance of Attentive FP compared to GATs. This is shown in Figure 6.

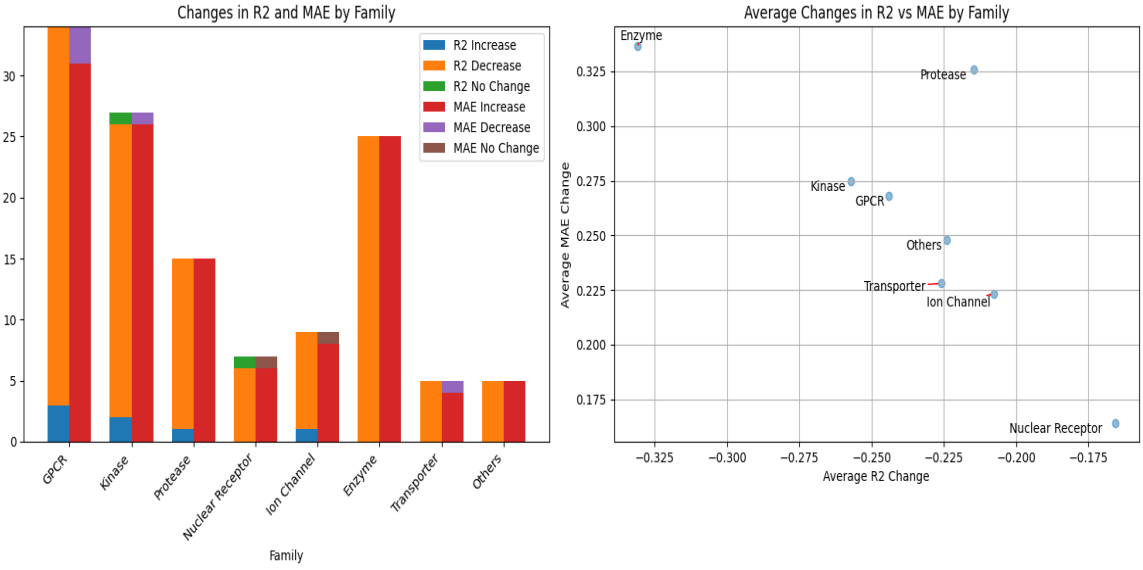

Figure 6: Before and after truncation performance of GATs

Similarly, a few targets in the GPCR and Kinase families saw some perfor-

mance gains. Not only GATs appeared to be more resistant to changes relative to Attentive FP, the average changes in MAE and R^2 by family also appeared more predictable. This implies that GATs may be a better choice for binding affinity prediction tasks when data is limited. The full results for the main and truncated datasets can be found in Table 13 and 15.

5.3.2 Attentive FP

For Attentive FP, the results were mostly similar, but there were a few that improved. We show this in Figure 7, where we graph both the number of changes between MAE and R^2, and also the average changes in MAE and R^2 by family.

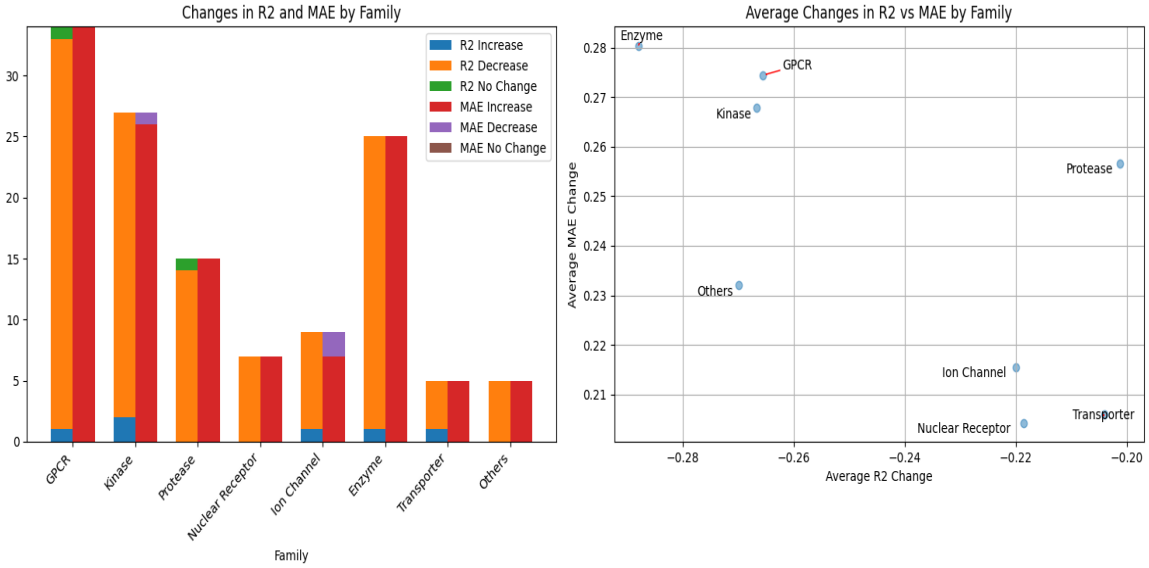

Figure 7: Before and after truncation performance of Attentive FP

Out of the test targets. The Enzyme family appeared most prone to changes, while Transporters, Ion Channels, and Nuclear Receptors were the most resilient. The full results for the main and truncated datasets can be found in Table 14 and 16.

Chapter VI.

Conclusion

6.1. Summary

This thesis evaluated the effectiveness of GATs and Attentive FP in comparison to GCNs for predicting molecular compound activities. The results indicate that GATs and Attentive FP perform comparably to GCNs in predicting the binding affinity of molecular targets, showcasing their utility in molecular simulations. We observed that the Protease and Transporter families appear to be more resilient to changes compared to other families when we truncate our datasets. An interesting observation from our study is how different targets reacted to dataset truncation. Specifically, GABA Transporter 1 showed improved performance with GCNs after truncation. In contrast, ADAM17 maintained consistent performance across all models despite a data reduction exceeding 80%, showing its robustness.

This research paves the way to further explore the potential of GATs and Attentive FP networks in drug discovery and molecular property prediction. We note that the choice of model should be made based on the dataset and the family of the target protein. We also highlight the importance of using a broad spectrum

of chemical diversity to ensure accurate generalization to unknown molecules. We believe these findings will inspire and guide future efforts in the fields of molecular property prediction and drug discovery.

6.2. Future Directions

Building upon the strong foundation laid by this thesis, future research could explore the performance of models on 3D molecular structures to potentially capture more intricate geometric details that are not represented in 2D SMILES sequences. Moon et al. (2023) demonstrates that leveraging 3D positions enables the utilization of unique molecular geometric properties such as distance, angle, and torsion in Euclidean space. This may lead to more accurate predictions of molecular properties.

To further improve the robustness of predictive models, it is recommended that future studies expand the diversity of data sources. While the datasets employed in this thesis provided valuable insights, a more comprehensive dataset that includes a wider array of protein families and data points would be beneficial. This could be complemented by integrating additional data types such as protein sequences and 3D structural information to enable a more holistic analysis.

Finally, the use of large synthetic datasets should be explored. Synthetic data can supplement training, especially when empirical data is limited. Volgin et al. (2022) demonstrated the effectiveness of synthetic data in model training, suggesting this approach could also enhance molecular property prediction models. Implement-

ing synthetic datasets could be particularly valuable in drug discovery when empirical data is not available. These proposed directions not only aim to refine the predictive accuracy of molecular property models but also hold the potential to significantly accelerate the pace of drug discovery and development.

References

Baldi, P. & Chauvin, Y. (1996). Hybrid modeling, hmm/nn architectures, and protein applications. *Neural computation*, 8, 1541–65.

Bosc, N., Atkinson, F., Félix, E., Gaulton, A., Hersey, A., & Leach, A. (2019). Large scale comparison of qsar and conformal prediction methods and their applications in drug discovery. *Journal of Cheminformatics*, 11, 4.

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug Discovery Today*, 23(6), 1241–1250.

Chen, J., Si, Y. W., Un, C.-W., & Siu, S. (2021). Chemical toxicity prediction based on semi-supervised learning and graph convolutional neural network. *Journal of Cheminformatics*, 13.

Cremer, J., Medrano Sandonas, L., Tkatchenko, A., Clevert, D.-A., & Fabritiis, G. (2023). Equivariant graph neural networks for toxicity prediction.

Dejun, J., Wu, Z., Hsieh, K., Guangyong, C., Liao, B., Wang, Z., Shen, C., Cao, D.-S., Wu, J., & Hou, T. (2020). Could graph neural networks learn better molecular

representation for drug discovery? a comparison study of descriptor-based and graph-based models.

Fung, V., Zhang, J., Juarez, E., & Sumpter, B. (2021). Benchmarking graph neural networks for materials chemistry. *npj Computational Materials*, 7.

Gawehn, E., Hiss, J. A., & Schneider, G. (2016). 6 deep learning in drug discovery.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry.

Gori, M., Monfardini, G., & Scarselli, F. (2005). A new model for learning in graph domains. volume 2 (pp. 729 – 734 vol. 2).

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362.

Hu, P., Jiao, Z., Zhang, Z., & Wang, Q. (2021). Development of solubility prediction models with ensemble learning. *Industrial & Engineering Chemistry Research*, 60(30), 11627–11635.

Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2, 573–584.

Jiménez-Luna, J. & Ginebra, J. (2017). pygpgo: Bayesian optimization for python. *The Journal of Open Source Software*, 2, 431.

Ketkar, R., Liu, Y., Wang, H., & Tian, H. (2023). A benchmark study of graph models for molecular acute toxicity prediction. *International Journal of Molecular Sciences*, 24, 11966.

Kipf, T. N. & Welling, M. (2017). Semi-supervised classification with graph convolutional networks.

Kojima, R., Ishida, S., Ohta, M., Iwata, H., Honma, T., & Okuno, Y. (2020). kGCN: a graph-based deep learning framework for chemical structures. *Journal of Cheminformatics*, 12(1).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86, 2278 – 2324.

Lee, S., Park, H., Choi, C., Kim, W., Kim, K., Han, Y.-K., Kang, J., Kang, C.-J., & Son, Y. (2023). Multi-order graph attention network for water solubility prediction and interpretation. *Scientific Reports*, 13.

Lin, W. (2020). *Social Media Analytics with Graph Convolutional Networks*. PhD thesis, University of Toronto.

Liu, X., Ding, Z., Li, N., Chen, Y., & Zhao, D. (2021). Egcn: Ensemble graph

convolutional network for neural architecture performance prediction. In *2021 8th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)* (pp. 149–154).

Lu, C., Liu, Q., Wang, C., Huang, Z., Lin, P., & He, L. (2019). Molecular property prediction: A multilevel quantum interactions modeling perspective. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 1052–1060.

Lusci, A., Pollastri, G., & Baldi, P. (2013). Deep architectures and deep learning in chemoinformatics: The prediction of aqueous solubility for drug-like molecules. *Journal of chemical information and modeling*, 53.

McCardle, K. (2023). Shedding light on gnn affinity predictions. *Nat Comput Sci*, 3(12), 1004.

Mohammed, A. & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, 35.

Moon, K., Im, H.-J., & Kwon, S. (2023). 3D graph contrastive learning for molecular property prediction. *Bioinformatics*, 39(6), btad371.

pandas development team, T. (2020). pandas-dev/pandas: Pandas.

Perkins, R., Fang, H., Tong, W., & Welsh, W. (2003). Quantitative structure-activity relationship methods: Perspectives on drug discovery and toxicology. *Environmental toxicology and chemistry / SETAC*, 22, 1666–79.

Ramsundar, B., Eastman, P., Walters, P., Pande, V., Leswing, K., & Wu, Z. (2019). *Deep Learning for the Life Sciences*. O'Reilly Media.

Sakai, M., Nagayasu, K., Shibui, N., Andoh, C., Takayama, K., Shirakawa, H., & Kaneko, S. (2021). Prediction of pharmacological activities from chemical structures with graph convolutional neural networks. *Scientific Reports*, 11.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1), 61–80.

Shayanfar, A., Fakhree, M., & Jouyban, A. (2010). A simple qspr model to predict aqueous solubility of drugs. *Journal of Drug Delivery Science and Technology*, 20(6), 467–476.

Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., et al. (2020). A deep learning approach to antibiotic discovery. *Cell*, 180(4), 688–702.

Sun, J., Wen, M., Wang, H., Ruan, Y., Yang, Q., Kang, X., Zhang, H., Zhang, Z., & Lu, H. (2022). Prediction of drug-likeness using graph convolutional attention network. *Bioinformatics*, 38.

Sun, M., Zhao, S., Gilvary, C., Elemento, O., Zhou, J., & Wang, F. (2020). Graph convolutional networks for computational drug development and discovery. *Briefings in bioinformatics*, 21(3), 919–935.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks.

Volgin, I., Batyr, P., Matseevich, A., Dobrovskiy, A., Andreeva, M., Nazarychev, V., Larin, S., Goikhman, M., Vizilter, Y., Askadskii, A., & Lyulin, S. (2022). Machine learning with enormous "synthetic" data sets: Predicting glass transition temperature of polyimides using graph convolutional neural networks. *ACS Omega*, 7.

Wang, H., Jiang, Z., You, Y., Han, Y., Liu, G., Srinivasa, J., Kompella, R. R., & Wang, Z. (2023). Graph mixture of experts: Learning on large-scale graphs with explicit diversity modeling.

Wu, L., Sun, P., Hong, R., Fu, Y., Wang, X., & Wang, M. (2019). Socialgcn: An efficient graph convolutional network based model for social recommendation.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24.

Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., Leswing, K., & Pande, V. (2018). Moleculenet: a benchmark for molecular machine learning. *Chem. Sci.*, 9, 513–530.

Wu, Z., Wang, J., Du, H., dejun, J., Kang, Y., Li, D., Pan, P., Deng, Y., Cao, D.-S., Hsieh, K., & Hou, T. (2023). Chemistry-intuitive explanation of graph neural

networks for molecular property prediction with substructure masking. *Nature Communications*, 14.

Xiong, Z., Wang, D., Liu, X., Feisheng, Z., Wan, X., Li, X., Li, Z., Luo, X., Chen, K., Jiang, H., & Zheng, M. (2019). Pushing the boundaries of molecular representation for drug discovery with graph attention mechanism. *Journal of Medicinal Chemistry*, 63.

Yang, K., Swanson, K., Jin, W., Coley, C., Eiden, P., Gao, H., Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., Palmer, A., Settels, V., Jaakkola, T., Jensen, K., & Barzilay, R. (2019). Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59.

Ying, R., Bourgeois, D., You, J., Zitnik, M., & Leskovec, J. (2019). Gnnexplainer: Generating explanations for graph neural networks.

Zhang, J., Gao, C., Jin, D., & Li, Y. (2020). Group-buying recommendation for social e-commerce.

Zhang, X.-M., Liang, L., Liu, L., & Tang, M. (2021). Graph neural networks and their current applications in bioinformatics. *Frontiers in Genetics*, 12.

Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., Terentiev, V. A., Polykovskiy, D. A., Kuznetsov, M. D., Asadulaev, A., Volkov, Y., Zholus, A., Shayakhmetov, R., Zhebrak, A., Minaeva, L. I.,

Zagribelnyy, B. A., Lee, L. H., Soll, R., Madge, D., Xing, L., Guo, T., Aspuru-Guzik, A., Winkler, D. A., & Makarov, A. A. (2019). Deep learning enables rapid identification of potent ddr1 kinase inhibitors. *Nature Biotechnology*, 37, 1038–1040.

Zhu, W., Zhang, Y., Zhao, D., Xu, J., & Wang, L. (2022). Hignn: Hierarchical informative graph neural networks for molecular property prediction equipped with feature-wise attention.

Appendix A.

Source code for this work is available upon request on Github.

Appendix B.

```
def hyper_model(params, train_dataset, valid_dataset):
    model = create_model(params, train_dataset, valid_dataset)
    valid_scores = model.evaluate(valid_dataset, [Metric(
 mean_absolute_error), Metric(pearson_r2_score)], transformers
 =[])
    return -valid_scores['mean_absolute_error'] + 2 *
 valid_scores['pearson_r2_score']

...
    params_dict = {
        'graph_attention_layer_size': ('int', [32, 2048]),
        'n_attention_heads': ('int', [2, 16]),
        'dropout': ('cont', [0.0, 0.5]),
        'alpha': ('cont', [0.0, 1.0]),
        'predictor_hidden_feats': ('int', [32, 512]),
        'predictor_dropout': ('cont', [0.0, 0.5]),
        'learning_rate': ('cont', [0.0001, 0.0020]),
        'weight_decay': ('cont', [0.0, 0.01])
    }
    cov = matern32()
    gp = GaussianProcess(cov)
    acq = Acquisition(mode='ExpectedImprovement')

    gpgo = GPGO(gp, acq, lambda **params: hyper_model(params,
 train_dataset, valid_dataset), params_dict)
    gpgo.run(max_iter=10, init_evals=1)
    best_params = gpgo.getResult()[0]
    save_hyperparameters(best_params)
```

```
 def hyper_model(params, train_dataset, valid_dataset):
     model = create_model(params, train_dataset, valid_dataset
)
     valid_scores = model.evaluate(valid_dataset, [Metric(
mean_absolute_error), Metric(pearson_r2_score)], transformers
=[])
     return -valid_scores['mean_absolute_error'] + 2 *
valid_scores['pearson_r2_score']

 ...
     params_dict = {
         'learning_rate': ('cont', [0.0001, 0.002]),
         'num_layers': ('int', [1, 5]),
         'num_timesteps': ('int', [1, 5]),
         'graph_feat_size': ('int', [100, 300]),
         'dropout': ('cont', [0.0, 0.5])
     }
     cov = matern32()
     gp = GaussianProcess(cov)
     acq = Acquisition(mode='ExpectedImprovement')

     gpgo = GPGO(gp, acq, lambda **params: hyper_model(params,
 train_dataset, valid_dataset), params_dict)
     gpgo.run(max_iter=10, init_evals=1)
     best_params = gpgo.getResult()[0]
     save_hyperparameters(best_params)
```

Appendix D.

Table 13: GAT Performance for 127 Targets

Protein Family	Target	MAE	R^2
GPCR	Orexin Receptor 1	0.475512	0.636221
GPCR	Corticotropin Releasing Factor Receptor 1	0.515567	0.65453
GPCR	Orexin Receptor 2	0.526359	0.612372
GPCR	Adenosine A2A Receptor	0.551854	0.678465
GPCR	Dopamine D3 Receptor	0.558624	0.636672
GPCR	Serotonin 6 5-HT6 Receptor	0.563044	0.585901
GPCR	Dopamine D2 Receptor	0.564776	0.541897
GPCR	Dopamine D1 Receptor	0.578766	0.559504
GPCR	Metabotropic Glutamate Receptor 5	0.587709	0.613008
GPCR	Melanin-Concentrating Hormone Receptor 1	0.589468	0.406779
GPCR	Histamine H3 Receptor	0.596571	0.749566
GPCR	Cannabinoid CB1 Receptor	0.600483	0.578084
GPCR	G Protein-Coupled Receptor 44	0.607457	0.559104
GPCR	Adenosine A1 Receptor	0.613453	0.556764
GPCR	Serotonin 2A 5-HT2A Receptor	0.614568	0.603278
GPCR	Cannabinoid CB2 Receptor	0.617463	0.594979
GPCR	Adenosine A3 Receptor	0.621353	0.613821
GPCR	Serotonin 7 5-HT7 Receptor	0.62795	0.528161
GPCR	Kappa Opioid Receptor	0.628353	0.693432
GPCR	Delta Opioid Receptor	0.628962	0.679433
GPCR	Neurokinin 1 Receptor	0.636539	0.67147
GPCR	Sigma Opioid Receptor	0.654721	0.441043
GPCR	Serotonin 1A 5-HT1A Receptor	0.665991	0.554044
GPCR	Muscarinic Acetylcholine Receptor M1	0.688145	0.767622
GPCR	Mu Opioid Receptor	0.689009	0.645853
GPCR	Cholecystokinin A Receptor	0.690741	0.522756
GPCR	Cholecystokinin B Receptor	0.703806	0.567417
GPCR	Dopamine D4 Receptor	0.704964	0.398064
GPCR	Melanocortin Receptor 4	0.718132	0.465346

GPCR	Muscarinic Acetylcholine Receptor M2	0.753243	0.751521
GPCR	Alpha-1A Adrenergic Receptor	0.789776	0.707303
GPCR	Endothelin Receptor ET-A	0.798292	0.516032
GPCR	Gonadotropin-Releasing Hormone Receptor	0.807326	0.86102
GPCR	Serotonin 2c 5-HT2C Receptor	0.825965	0.585749
Enzyme	Acetyl-CoA Carboxylase 2	0.348305	0.59923
Enzyme	Nicotinamide Phosphoribosyltransferase	0.431985	0.623034
Enzyme	Poly ADP-Ribose Polymerase-1	0.475141	0.737627
Enzyme	Protein-Tyrosine Phosphatase 1B	0.494986	0.656528
Enzyme	Cyclooxygenase-1	0.509323	0.419991
Enzyme	Arachidonate 5-Lipoxygenase	0.5173	0.502931
Enzyme	Carbonic Anhydrase I	0.530313	0.680667
Enzyme	11-Beta-Hydroxysteroid Dehydrogenase 1	0.561321	0.732221
Enzyme	Cholinesterase	0.563629	0.773046
Enzyme	Butyrylcholinesterase	0.563629	0.773046
Enzyme	Monoamine Oxidase A	0.566847	0.668368
Enzyme	Carbonic Anhydrase IX	0.589344	0.592715
Enzyme	Carbonic Anhydrase II	0.591008	0.638983
Enzyme	Phosphodiesterase 10A	0.600313	0.613031
Enzyme	Integrase	0.604866	0.718087
Enzyme	Carbonic Anhydrase XII	0.610001	0.612983
Enzyme	Cytochrome P450 19A1	0.616699	0.630775
Enzyme	Human Immunodeficiency Virus Type 1 Reverse Transcriptase	0.631387	0.596076
Enzyme	Acetylcholinesterase	0.637671	0.65907
Enzyme	Cyclooxygenase-2	0.665506	0.573248
Enzyme	Monoamine Oxidase B	0.683699	0.63055
Enzyme	Anandamide Amidohydrolase	0.715577	0.802681
Enzyme	Dihydrofolate Reductase	0.722961	0.587692
Enzyme	Protein Farnesyltransferase	0.771041	0.650968
Enzyme	Coagulation Factor X	0.82506	0.651424
Ion Channel	hERG	0.477247	0.524965
Ion Channel	Sodium Channel Protein Type IX Alpha Subunit	0.481973	0.648225
Ion Channel	Voltage-Gated Potassium Channel Subunit Kv1.5	0.497179	0.458191
Ion Channel	Transient Receptor Potential Cation Channel Subfamily M Member 8	0.519204	0.828398
Ion Channel	Vanilloid Receptor	0.541078	0.684196

Ion Channel	P2X Purinoceptor 7	0.599509	0.374953
Ion Channel	Transient Receptor Potential Cation Channel Subfamily A Member 1	0.641016	0.502374
Ion Channel	Neuronal Acetylcholine Receptor Protein Alpha-7 Subunit	0.661421	0.627767
Ion Channel	Neuronal Acetylcholine Receptor Alpha4 Beta2	0.949944	0.533172
Kinase	Fibroblast Growth Factor Receptor 3	0.447973	0.706114
Kinase	Nerve Growth Factor Receptor Trk-A	0.471246	0.616584
Kinase	Serine Threonine-Protein Kinase mTOR	0.478746	0.752173
Kinase	Serine Threonine-Protein Kinase B-Raf	0.488564	0.734272
Kinase	Pi3-Kinase P110-Alpha Subunit	0.505521	0.705926
Kinase	Insulin-Like Growth Factor I Receptor	0.509064	0.784112
Kinase	Serine Threonine-Protein Kinase Pim1	0.522521	0.823863
Kinase	Hepatocyte Growth Factor Receptor	0.532805	0.703616
Kinase	Fibroblast Growth Factor Receptor 1	0.541861	0.759191
Kinase	Tyrosine-Protein Kinase JAK1	0.550293	0.636757
Kinase	Serine Threonine-Protein Kinase Pim2	0.557517	0.681565
Kinase	Serine Threonine-Protein Kinase Akt	0.569934	0.692662
Kinase	Tyrosine-Protein Kinase SYK	0.571447	0.645326
Kinase	Map Kinase ERK2	0.571936	0.769534
Kinase	Tyrosine-Protein Kinase JAK2	0.578351	0.67511
Kinase	Epidermal Growth Factor Receptor ErbB1	0.578622	0.713641
Kinase	PI3-Kinase P110-Delta Subunit	0.582241	0.679293
Kinase	Vascular Endothelial Growth Factor Receptor 2	0.587682	0.62532
Kinase	Tyrosine-Protein Kinase Receptor Flt3	0.595912	0.612245
Kinase	Map Kinase P38 Alpha	0.602107	0.679678
Kinase	PI3-Kinase P110-Gamma Subunit	0.602945	0.589654
Kinase	Tyrosine-Protein Kinase JAK3	0.614309	0.764549
Kinase	Cyclin-Dependent Kinase 2	0.618524	0.605985
Kinase	Tyrosine-Protein Kinase ABL	0.626921	0.781662
Kinase	Tyrosine-Protein Kinase Src	0.647486	0.690192
Kinase	Serine Threonine-Protein Kinase Aurora-A	0.660356	0.682692
Kinase	Glycogen Synthase Kinase-3 Beta	0.766983	0.446199
Nuclear Receptor	Thyroid Hormone Receptor Alpha	0.432221	0.838262
Nuclear Receptor	Androgen Receptor	0.509433	0.741759

Nuclear Receptor	Vitamin D Receptor	0.552683	0.819698
Nuclear Receptor	Peroxisome Proliferator-Activated Receptor Gamma	0.573776	0.680704
Nuclear Receptor	Glucocorticoid Receptor	0.606133	0.650094
Nuclear Receptor	Estrogen Receptor Alpha	0.657432	0.725885
Nuclear Receptor	Estrogen Receptor Beta	0.755918	0.655699
Protease	Matrix Metalloproteinase-1	0.468225	0.652845
Protease	Cathepsin D	0.501263	0.804864
Protease	Adam17	0.510525	0.849666
Protease	Beta-Secretase 1	0.590782	0.635515
Protease	Matrix Metalloproteinase 9	0.594595	0.710221
Protease	Cathepsin S	0.604124	0.689075
Protease	Dipeptidyl Peptidase IV	0.609874	0.723071
Protease	Matrix Metalloproteinase 13	0.635367	0.684301
Protease	Leukocyte Elastase	0.638669	0.923682
Protease	Gamma-Secretase	0.642171	0.638409
Protease	Matrix Metalloproteinase-2	0.642226	0.763534
Protease	Thrombin	0.65373	0.730604
Protease	Renin	0.82019	0.580784
Protease	Trypsin I	0.866903	0.631662
Protease	Human Immunodeficiency Virus Type 1 Protease	0.964416	0.491982
Transporter	GABA Transporter 1	0.495396	0.743078
Transporter	Potassium-Transporting ATPase	0.49603	0.587555
Transporter	Dopamine Transporter	0.55616	0.737176
Transporter	Serotonin Transporter	0.566232	0.685962
Transporter	Norepinephrine Transporter	0.618925	0.53601
Others	P53-Binding Protein MDM-2	0.466591	0.883679
Others	Histone Deacetylase 1	0.497822	0.670272
Others	Histone Deacetylase 6	0.522767	0.734411
Others	Bromodomain-Containing Protein 4	0.528409	0.710369
Others	Apoptosis Regulator BCL-2	0.542345	0.854652

Appendix E.

Table 14: Attentive FP Performance for 127 Targets

Protein Family	Target	MAE	R^2
GPCR	Corticotropin Releasing Factor Receptor 1	0.490922	0.623825
GPCR	Orexin Receptor 1	0.497706	0.591447
GPCR	Dopamine D3 Receptor	0.531025	0.663432
GPCR	Serotonin 6 5-HT6 Receptor	0.556455	0.612263
GPCR	Dopamine D1 Receptor	0.564368	0.577392
GPCR	Dopamine D2 Receptor	0.570426	0.521134
GPCR	Histamine H3 Receptor	0.576515	0.748236
GPCR	Kappa Opioid Receptor	0.578933	0.712382
GPCR	Melanin-Concentrating Hormone Receptor 1	0.583885	0.412057
GPCR	G Protein-Coupled Receptor 44	0.586915	0.526013
GPCR	Cannabinoid CB1 Receptor	0.588262	0.604333
GPCR	Adenosine A2A Receptor	0.590046	0.619552
GPCR	Orexin Receptor 2	0.597953	0.518812
GPCR	Adenosine A3 Receptor	0.605214	0.614513
GPCR	Sigma Opioid Receptor	0.61019	0.529517
GPCR	Delta Opioid Receptor	0.611888	0.691807
GPCR	Serotonin 7 5-HT7 Receptor	0.626231	0.565231
GPCR	Serotonin 1A 5-HT1A Receptor	0.629163	0.571075
GPCR	Cholecystokinin A Receptor	0.633949	0.545477
GPCR	Metabotropic Glutamate Receptor 5	0.636199	0.576591
GPCR	Cannabinoid CB2 Receptor	0.638329	0.591378
GPCR	Serotonin 2A 5-HT2A Receptor	0.638762	0.534619
GPCR	Neurokinin 1 Receptor	0.644041	0.68635
GPCR	Adenosine A1 Receptor	0.648008	0.527609
GPCR	Muscarinic Acetylcholine Receptor M1	0.6569	0.815135
GPCR	Dopamine D4 Receptor	0.658841	0.48456
GPCR	Muscarinic Acetylcholine Receptor M2	0.666259	0.761169
GPCR	Serotonin 2C 5-HT2C Receptor	0.679467	0.607908
GPCR	Melanocortin Receptor 4	0.687404	0.499235

GPCR	Cholecystokinin B Receptor	0.69641	0.594627
GPCR	Mu Opioid Receptor	0.728871	0.570209
GPCR	Endothelin Receptor ET-A	0.738405	0.59112
GPCR	Gonadotropin-Releasing Hormone Receptor	0.752632	0.863116
GPCR	Alpha-1A Adrenergic Receptor	0.813947	0.676316
Enzyme	Acetyl-CoA Carboxylase 2	0.339897	0.59946
Enzyme	Nicotinamide Phosphoribosyltransferase	0.460304	0.552548
Enzyme	Poly ADP-Ribose Polymerase-1	0.477657	0.72486
Enzyme	Arachidonate 5-Lipoxygenase	0.506763	0.53998
Enzyme	Cyclooxygenase-1	0.516055	0.384046
Enzyme	Carbonic Anhydrase IX	0.516832	0.64339
Enzyme	Integrase	0.520679	0.774513
Enzyme	11-Beta-Hydroxysteroid Dehydrogenase 1	0.523863	0.754812
Enzyme	Carbonic Anhydrase XII	0.549979	0.642072
Enzyme	Protein-Tyrosine Phosphatase 1B	0.557424	0.605356
Enzyme	Butyrylcholinesterase	0.562489	0.78952
Enzyme	Cholinesterase	0.562489	0.78952
Enzyme	Carbonic Anhydrase I	0.569368	0.649002
Enzyme	Cytochrome P450 19A1	0.60352	0.635757
Enzyme	Acetylcholinesterase	0.61918	0.695379
Enzyme	Monoamine Oxidase A	0.621665	0.608386
Enzyme	Phosphodiesterase 10A	0.632917	0.599676
Enzyme	Human Immunodeficiency Virus Type 1 Reverse Transcriptase	0.666502	0.566929
Enzyme	Cyclooxygenase-2	0.755628	0.494175
Enzyme	Monoamine Oxidase B	0.756907	0.585437
Enzyme	Anandamide Amidohydrolase	0.781291	0.750104
Enzyme	Protein Farnesyltransferase	0.78212	0.60475
Enzyme	Dihydrofolate Reductase	0.808603	0.511523
Enzyme	Carbonic Anhydrase II	0.830163	0.500335
Enzyme	Coagulation Factor X	0.841557	0.649621
Ion Channel	hERG	0.402358	0.585391
Ion Channel	Voltage-Gated Potassium Channel Subunit Kv1.5	0.488535	0.430439
Ion Channel	Sodium Channel Protein Type IX Alpha Subunit	0.555087	0.577707
Ion Channel	Vanilloid Receptor	0.5731	0.609582
Ion Channel	P2X Purinoceptor 7	0.586491	0.360953
Ion Channel	Transient Receptor Potential Cation Channel Subfamily M Member 8	0.633245	0.732009

Ion Channel	Neuronal Acetylcholine Receptor Protein Alpha-7 Subunit	0.644547	0.661759
Ion Channel	Transient Receptor Potential Cation Channel Subfamily A Member 1	0.718269	0.494236
Ion Channel	Neuronal Acetylcholine Receptor Alpha4 Beta2	1.10829	0.543367
Kinase	Fibroblast Growth Factor Receptor 3	0.447547	0.681579
Kinase	Nerve Growth Factor Receptor Trk-A	0.44961	0.650777
Kinase	Insulin-Like Growth Factor I Receptor	0.457185	0.811463
Kinase	Fibroblast Growth Factor Receptor 1	0.458142	0.823691
Kinase	Serine Threonine-Protein Kinase B-Raf	0.47366	0.765909
Kinase	Serine Threonine-Protein Kinase mTOR	0.48659	0.751609
Kinase	Tyrosine-Protein Kinase SYK	0.508856	0.695146
Kinase	Tyrosine-Protein Kinase JAK1	0.514207	0.657184
Kinase	Serine Threonine-Protein Kinase Pim1	0.537158	0.81245
Kinase	Tyrosine-Protein Kinase JAK2	0.542957	0.680993
Kinase	PI3-Kinase P110-Gamma Subunit	0.555766	0.587181
Kinase	Tyrosine-Protein Kinase Receptor Flt3	0.556648	0.618616
Kinase	Map Kinase ERK2	0.562201	0.780079
Kinase	Serine Threonine-Protein Kinase Pim2	0.566125	0.681828
Kinase	PI3-Kinase P110-Delta Subunit	0.577136	0.692287
Kinase	Tyrosine-Protein Kinase JAK3	0.578918	0.765885
Kinase	Serine Threonine-Protein Kinase Akt	0.581336	0.67595
Kinase	Tyrosine-Protein Kinase ABL	0.591758	0.772936
Kinase	Vascular Endothelial Growth Factor Receptor 2	0.600296	0.639522
Kinase	Hepatocyte Growth Factor Receptor	0.607426	0.588303
Kinase	Map Kinase P38 Alpha	0.609208	0.657379
Kinase	Epidermal Growth Factor Receptor ErbB1	0.627375	0.692253
Kinase	Tyrosine-Protein Kinase SRC	0.62939	0.694305
Kinase	PI3-Kinase P110-Alpha Subunit	0.644473	0.57791
Kinase	Cyclin-Dependent Kinase 2	0.645197	0.586916
Kinase	Serine Threonine-Protein Kinase Aurora-A	0.711986	0.625632
Kinase	Glycogen Synthase Kinase-3 Beta	0.715222	0.518864
Nuclear Receptor	Thyroid Hormone Receptor Alpha	0.317999	0.905391
Nuclear Receptor	Vitamin D Receptor	0.501735	0.853711
Nuclear Receptor	Androgen Receptor	0.572326	0.676986

Nuclear Receptor	Peroxisome Proliferator-Activated Receptor Gamma	0.587907	0.665589
Nuclear Receptor	Glucocorticoid Receptor	0.60594	0.647011
Nuclear Receptor	Estrogen Receptor Alpha	0.670149	0.683154
Nuclear Receptor	Estrogen Receptor Beta	0.706901	0.65642
Protease	Matrix Metalloproteinase-1	0.455911	0.63966
Protease	ADAM17	0.457197	0.861113
Protease	Cathepsin D	0.501099	0.821792
Protease	Leukocyte Elastase	0.567666	0.946549
Protease	Matrix Metalloproteinase 9	0.584676	0.693282
Protease	Beta-Secretase 1	0.594107	0.636789
Protease	Cathepsin S	0.597208	0.681064
Protease	Matrix Metalloproteinase-2	0.604098	0.770955
Protease	Matrix Metalloproteinase 13	0.607412	0.679033
Protease	Dipeptidyl Peptidase IV	0.626407	0.710055
Protease	Thrombin	0.639469	0.738645
Protease	Gamma-Secretase	0.659384	0.576
Protease	Renin	0.768762	0.56038
Protease	Trypsin I	0.989373	0.621731
Protease	Human Immunodeficiency Virus Type 1 Protease	1.07298	0.49152
Transporter	GABA Transporter 1	0.477037	0.783785
Transporter	Potassium-Transporting ATPase	0.518178	0.546349
Transporter	Serotonin Transporter	0.541849	0.710211
Transporter	Norepinephrine Transporter	0.631822	0.544152
Transporter	Dopamine Transporter	0.670739	0.656383
Others	P53-Binding Protein MDM-2	0.48738	0.866853
Others	Histone Deacetylase 1	0.48994	0.659453
Others	Bromodomain-Containing Protein 4	0.53122	0.694395
Others	Histone Deacetylase 6	0.564483	0.711331
Others	Apoptosis Regulator BCL-2	0.568045	0.829819

Appendix F.

Table 15: GAT Performance for 127 Trimmed Targets

Protein Family	Target	MAE	R^2
GPCR	Orexin Receptor 2	0.589704	0.452898
GPCR	Dopamine D1 Receptor	0.60931	0.486423
GPCR	Orexin Receptor 1	0.612161	0.580473
GPCR	Cholecystokinin B Receptor	0.668391	0.694264
GPCR	G Protein-Coupled Receptor 44	0.671468	0.504582
GPCR	Corticotropin Releasing Factor Receptor 1	0.692644	0.405586
GPCR	Serotonin 2C 5-HT2C Receptor	0.70269	0.482913
GPCR	Melanin-Concentrating Hormone Receptor 1	0.720804	0.264373
GPCR	Metabotropic Glutamate Receptor 5	0.736192	0.521537
GPCR	Endothelin Receptor ET-A	0.73648	0.55544
GPCR	Sigma Opioid Receptor	0.762185	0.317478
GPCR	Serotonin 6 5-HT6 Receptor	0.836981	0.216342
GPCR	Muscarinic Acetylcholine Receptor M2	0.837103	0.58993
GPCR	Dopamine D4 Receptor	0.838033	0.234377
GPCR	Dopamine D3 Receptor	0.874233	0.272097
GPCR	Histamine H3 Receptor	0.899806	0.232627
GPCR	Cannabinoid CB2 Receptor	0.913493	0.329095
GPCR	Cholecystokinin A Receptor	0.928256	0.196413
GPCR	Adenosine A2A Receptor	0.934012	0.315788
GPCR	Dopamine D2 Receptor	0.943521	0.14754
GPCR	Delta Opioid Receptor	0.949053	0.578616
GPCR	Adenosine A3 Receptor	0.966615	0.154715
GPCR	Cannabinoid CB1 Receptor	0.968799	0.337417
GPCR	Serotonin 7 5-HT7 Receptor	0.980402	0.198827
GPCR	Melanocortin Receptor 4	0.992796	0.189346
GPCR	Neurokinin 1 Receptor	0.998566	0.344464
GPCR	Adenosine A1 Receptor	1.00151	0.254665
GPCR	Serotonin 2A 5-HT2A Receptor	1.02808	0.111837
GPCR	Muscarinic Acetylcholine Receptor M1	1.04708	0.646855

GPCR	Serotonin 1A 5-HT1A Receptor	1.08584	0.170924
GPCR	Kappa Opioid Receptor	1.19955	0.142706
GPCR	Mu Opioid Receptor	1.3573	0.19351
GPCR	Alpha-1A Adrenergic Receptor	1.36432	0.21584
GPCR	Gonadotropin-Releasing Hormone Receptor	1.47773	0.867682
Enzyme	Acetyl-CoA Carboxylase 2	0.424239	0.233817
Enzyme	Nicotinamide Phosphoribosyltransferase	0.599874	0.208696
Enzyme	Carbonic Anhydrase XII	0.669231	0.422563
Enzyme	Poly ADP-Ribose Polymerase-1	0.742217	0.566515
Enzyme	Cyclooxygenase-1	0.74357	0.196585
Enzyme	11-Beta-Hydroxysteroid Dehydrogenase 1	0.81701	0.506097
Enzyme	Protein-Tyrosine Phosphatase 1B	0.822669	0.199528
Enzyme	Arachidonate 5-Lipoxygenase	0.835897	0.138413
Enzyme	Protein Farnesyltransferase	0.845638	0.50651
Enzyme	Integrase	0.86877	0.629392
Enzyme	Cytochrome P450 19A1	0.875263	0.146494
Enzyme	Monoamine Oxidase B	0.914769	0.423846
Enzyme	Butyrylcholinesterase	0.917313	0.533758
Enzyme	Cholinesterase	0.917313	0.533758
Enzyme	Carbonic Anhydrase II	0.936086	0.314345
Enzyme	Carbonic Anhydrase IX	1.00677	0.219184
Enzyme	Anandamide Amidohydrolase	1.01289	0.537192
Enzyme	Phosphodiesterase 10A	1.02204	0.0384146
Enzyme	Dihydrofolate Reductase	1.06824	0.267371
Enzyme	Acetylcholinesterase	1.07549	0.142763
Enzyme	Human Immunodeficiency Virus Type 1 Reverse Transcriptase	1.18144	0.141049
Enzyme	Coagulation Factor X	1.24094	0.295657
Enzyme	Carbonic Anhydrase I	1.2513	0.286707
Enzyme	Cyclooxygenase-2	1.26347	0.160837
Enzyme	Monoamine Oxidase A	1.2954	0.123172
Ion Channel	Sodium Channel Protein Type IX Alpha Subunit	0.642392	0.402947
Ion Channel	Transient Receptor Potential Cation Channel Subfamily A Member 1	0.642817	0.548105
Ion Channel	Voltage-Gated Potassium Channel Subunit Kv1.5	0.6506	0.401565
Ion Channel	P2X Purinoceptor 7	0.669577	0.218545
Ion Channel	Transient Receptor Potential Cation Channel Subfamily M Member 8	0.675741	0.726565

Ion Channel	hERG	0.771669	0.192845
Ion Channel	Vanilloid Receptor	0.826548	0.337059
Ion Channel	Neuronal Acetylcholine Receptor Protein Alpha-7 Subunit	0.931706	0.402711
Ion Channel	Neuronal Acetylcholine Receptor Alpha4 Beta2	1.57182	0.097581
Kinase	Tyrosine-Protein Kinase SYK	0.550226	0.501422
Kinase	Nerve Growth Factor Receptor Trk-A	0.562369	0.319242
Kinase	Serine Threonine-Protein Kinase B-Raf	0.679553	0.546901
Kinase	PI3-Kinase P110-Delta Subunit	0.698697	0.49173
Kinase	Fibroblast Growth Factor Receptor 3	0.707076	0.482198
Kinase	Tyrosine-Protein Kinase Src	0.727675	0.688911
Kinase	Serine Threonine-Protein Kinase Aurora-A	0.741423	0.688579
Kinase	Serine Threonine-Protein Kinase mTOR	0.741542	0.542931
Kinase	Hepatocyte Growth Factor Receptor	0.753035	0.481951
Kinase	Cyclin-Dependent Kinase 2	0.759393	0.45449
Kinase	Tyrosine-Protein Kinase Receptor Flt3	0.763853	0.490944
Kinase	Serine Threonine-Protein Kinase Pim2	0.780631	0.311695
Kinase	Tyrosine-Protein Kinase JAK2	0.798446	0.275404
Kinase	Tyrosine-Protein Kinase JAK1	0.810062	0.376184
Kinase	Serine Threonine-Protein Kinase Pim1	0.816236	0.710495
Kinase	PI3-Kinase P110-Alpha Subunit	0.871526	0.309746
Kinase	Map Kinase p38 Alpha	0.882311	0.228604
Kinase	Serine Threonine-Protein Kinase Akt	0.895098	0.527702
Kinase	Tyrosine-Protein Kinase JAK3	0.900458	0.793027
Kinase	Insulin-Like Growth Factor I Receptor	0.927926	0.487329
Kinase	PI3-Kinase P110-Gamma Subunit	0.978141	0.0642752
Kinase	Tyrosine-Protein Kinase ABL	0.99634	0.573457
Kinase	Map Kinase ERK2	1.02825	0.297892
Kinase	Glycogen Synthase Kinase-3 Beta	1.04363	0.29941
Kinase	Fibroblast Growth Factor Receptor 1	1.04414	0.345846
Kinase	Vascular Endothelial Growth Factor Receptor 2	1.12646	0.0718916
Kinase	Epidermal Growth Factor Receptor ErbB1	1.19671	0.244884
Nuclear Receptor	Thyroid Hormone Receptor Alpha	0.432221	0.838262
Nuclear Receptor	Vitamin D Receptor	0.676815	0.747461
Nuclear Receptor	Glucocorticoid Receptor	0.70666	0.562693

Nuclear Receptor	Estrogen Receptor Alpha	0.774727	0.600589
Nuclear Receptor	Estrogen Receptor Beta	0.835242	0.509369
Nuclear Receptor	Peroxisome Proliferator-Activated Receptor Gamma	0.849526	0.379332
Nuclear Receptor	Androgen Receptor	0.957716	0.316801
Protease	Cathepsin D	0.57719	0.67336
Protease	ADAM17	0.684251	0.803204
Protease	Matrix Metalloproteinase-1	0.720367	0.490284
Protease	Beta-Secretase 1	0.733293	0.357651
Protease	Cathepsin S	0.837661	0.474929
Protease	Gamma-Secretase	0.870358	0.512929
Protease	Dipeptidyl Peptidase IV	0.908789	0.351477
Protease	Renin	0.923925	0.422101
Protease	Trypsin I	0.966424	0.723052
Protease	Matrix Metalloproteinase 9	0.969627	0.521593
Protease	Thrombin	0.986152	0.505473
Protease	Matrix Metalloproteinase 13	1.04879	0.364442
Protease	Matrix Metalloproteinase-2	1.17358	0.328815
Protease	Human Immunodeficiency Virus Type 1 Protease	1.45177	0.254783
Protease	Leukocyte Elastase	1.78818	0.512567
Transporter	Potassium-Transporting ATPase	0.489583	0.553208
Transporter	GABA Transporter 1	0.539342	0.692818
Transporter	Norepinephrine Transporter	0.827091	0.4493
Transporter	Dopamine Transporter	0.97181	0.268033
Transporter	Serotonin Transporter	1.04551	0.192957
Others	Bromodomain-Containing Protein 4	0.605324	0.568015
Others	Apoptosis Regulator BCL-2	0.704223	0.75977
Others	Histone Deacetylase 6	0.727624	0.555039
Others	Histone Deacetylase 1	0.878685	0.233667
Others	P53-Binding Protein MDM-2	0.884695	0.611539

Appendix G.

Table 16: Attentive FP Performance for 127 Trimmed Targets

Protein Family	Target	MAE	R^2
GPCR	Dopamine D1 Receptor	0.682311	0.452283
GPCR	Orexin Receptor 2	0.700794	0.290899
GPCR	G Protein-Coupled Receptor 44	0.701444	0.388748
GPCR	Sigma Opioid Receptor	0.728259	0.298365
GPCR	Corticotropin Releasing Factor Receptor 1	0.732187	0.333108
GPCR	Metabotropic Glutamate Receptor 5	0.732705	0.447071
GPCR	Orexin Receptor 1	0.736025	0.493911
GPCR	Cholecystokinin B Receptor	0.754511	0.589814
GPCR	Melanin-Concentrating Hormone Receptor 1	0.78178	0.227961
GPCR	Serotonin 6 5-HT6 Receptor	0.782301	0.347761
GPCR	Adenosine A3 Receptor	0.79487	0.200585
GPCR	Serotonin 2C 5-HT2C Receptor	0.804422	0.238801
GPCR	Adenosine A2A Receptor	0.812463	0.352419
GPCR	Delta Opioid Receptor	0.826328	0.648678
GPCR	Cannabinoid CB2 Receptor	0.827391	0.379538
GPCR	Neurokinin 1 Receptor	0.83114	0.506436
GPCR	Cholecystokinin A Receptor	0.835886	0.198803
GPCR	Adenosine A1 Receptor	0.836361	0.412129
GPCR	Dopamine D4 Receptor	0.88283	0.199972
GPCR	Endothelin Receptor ET-A	0.929514	0.483295
GPCR	Histamine H3 Receptor	0.929878	0.214152
GPCR	Serotonin 1A 5-HT1A Receptor	0.936297	0.233051
GPCR	Dopamine D2 Receptor	0.975995	0.112275
GPCR	Serotonin 7 5-HT7 Receptor	0.983921	0.20695
GPCR	Serotonin 2A 5-HT2A Receptor	0.994333	0.0771701
GPCR	Cannabinoid CB1 Receptor	1.01057	0.318776
GPCR	Melanocortin Receptor 4	1.0136	0.231703
GPCR	Mu Opioid Receptor	1.01997	0.326846
GPCR	Dopamine D3 Receptor	1.04419	0.204324

GPCR	Muscarinic Acetylcholine Receptor M2	1.04455	0.42102
GPCR	Kappa Opioid Receptor	1.07605	0.268229
GPCR	Gonadotropin-Releasing Hormone Receptor	1.08573	0.938868
GPCR	Muscarinic Acetylcholine Receptor M1	1.13591	0.475863
GPCR	Alpha-1A Adrenergic Receptor	1.67084	0.0776665
Enzyme	Acetyl-CoA Carboxylase 2	0.517972	0.231402
Enzyme	Nicotinamide Phosphoribosyltransferase	0.647807	0.172251
Enzyme	Cyclooxygenase-1	0.663365	0.198229
Enzyme	Poly ADP-Ribose Polymerase-1	0.720633	0.547431
Enzyme	Arachidonate 5-Lipoxygenase	0.724115	0.263447
Enzyme	Cytochrome P450 19A1	0.740117	0.345085
Enzyme	Protein-Tyrosine Phosphatase 1B	0.748505	0.271287
Enzyme	Carbonic Anhydrase XII	0.751454	0.307675
Enzyme	Carbonic Anhydrase IX	0.756133	0.437756
Enzyme	Protein Farnesyltransferase	0.800804	0.527304
Enzyme	11-Beta-Hydroxysteroid Dehydrogenase 1	0.808475	0.407708
Enzyme	Integrase	0.86361	0.581748
Enzyme	Cyclooxygenase-2	0.918556	0.547687
Enzyme	Anandamide Amidohydrolase	0.964452	0.715723
Enzyme	Butyrylcholinesterase	0.981487	0.412026
Enzyme	Cholinesterase	0.981487	0.412026
Enzyme	Phosphodiesterase 10A	0.988912	0.0704144
Enzyme	Monoamine Oxidase A	0.992852	0.26462
Enzyme	Monoamine Oxidase B	1.06341	0.309571
Enzyme	Carbonic Anhydrase II	1.06829	0.203616
Enzyme	Acetylcholinesterase	1.06934	0.171395
Enzyme	Carbonic Anhydrase I	1.114	0.195443
Enzyme	Dihydrofolate Reductase	1.1177	0.233031
Enzyme	Human Immunodeficiency Virus Type 1 Reverse Transcriptase	1.17997	0.21327
Enzyme	Coagulation Factor X	1.46845	0.134097
Ion Channel	Voltage-Gated Potassium Channel Subunit Kv1.5	0.608806	0.371786
Ion Channel	Transient Receptor Potential Cation Channel Subfamily M Member 8	0.614203	0.784754
Ion Channel	Sodium Channel Protein Type IX Alpha Subunit	0.694951	0.329833
Ion Channel	Transient Receptor Potential Cation Channel Subfamily A Member 1	0.708502	0.396932

Ion Channel	P2X Purinoceptor 7	0.807999	0.0223141
Ion Channel	Neuronal Acetylcholine Receptor Protein Alpha-7 Subunit	0.823346	0.50744
Ion Channel	Vanilloid Receptor	0.922438	0.322762
Ion Channel	hERG	1.02747	0.0898554
Ion Channel	Neuronal Acetylcholine Receptor Alpha4 Beta2	1.44172	0.198465
Kinase	Serine Threonine-Protein Kinase Aurora-A	0.602185	0.744449
Kinase	Fibroblast Growth Factor Receptor 3	0.622104	0.544021
Kinase	Nerve Growth Factor Receptor Trk-A	0.63105	0.294513
Kinase	Tyrosine-Protein Kinase Receptor Flt3	0.686655	0.478522
Kinase	Tyrosine-Protein Kinase SYK	0.696014	0.386
Kinase	Tyrosine-Protein Kinase JAK3	0.696355	0.877686
Kinase	Map Kinase ERK2	0.712688	0.592067
Kinase	Tyrosine-Protein Kinase JAK2	0.727495	0.463981
Kinase	Serine Threonine-Protein Kinase B-Raf	0.776089	0.498727
Kinase	PI3-Kinase P110-Gamma Subunit	0.780153	0.151872
Kinase	Serine Threonine-Protein Kinase mTOR	0.787296	0.415848
Kinase	Tyrosine-Protein Kinase ABL	0.790516	0.700872
Kinase	Hepatocyte Growth Factor Receptor	0.829403	0.416103
Kinase	Map Kinase P38 Alpha	0.848987	0.208295
Kinase	Serine Threonine-Protein Kinase Pim2	0.855006	0.331146
Kinase	Serine Threonine-Protein Kinase Pim1	0.878225	0.658164
Kinase	Tyrosine-Protein Kinase SRC	0.885249	0.497705
Kinase	Serine Threonine-Protein Kinase Akt	0.887022	0.507695
Kinase	PI3-Kinase P110-Delta Subunit	0.898694	0.324156
Kinase	PI3-Kinase P110-Alpha Subunit	0.902458	0.209273
Kinase	Tyrosine-Protein Kinase JAK1	0.910165	0.33833
Kinase	Fibroblast Growth Factor Receptor 1	0.93935	0.368842
Kinase	Insulin-Like Growth Factor I Receptor	0.998174	0.357683
Kinase	Glycogen Synthase Kinase-3 Beta	1.01375	0.312612
Kinase	Epidermal Growth Factor Receptor ErbB1	1.02061	0.274344
Kinase	Vascular Endothelial Growth Factor Receptor 2	1.04763	0.130452
Kinase	Cyclin-Dependent Kinase 2	1.04828	0.225797
Nuclear Receptor	Thyroid Hormone Receptor Alpha	0.353245	0.875037
Nuclear Receptor	Vitamin D Receptor	0.718436	0.735657

Nuclear Receptor	Glucocorticoid Receptor	0.780833	0.423534
Nuclear Receptor	Estrogen Receptor Alpha	0.815322	0.467774
Nuclear Receptor	Peroxisome Proliferator-Activated Receptor Gamma	0.881489	0.291319
Nuclear Receptor	Estrogen Receptor Beta	0.897794	0.492477
Nuclear Receptor	Androgen Receptor	0.943881	0.270307
Protease	ADAM17	0.5552	0.812146
Protease	Matrix Metalloproteinase-1	0.69898	0.463085
Protease	Cathepsin D	0.722064	0.539881
Protease	Matrix Metalloproteinase 9	0.727469	0.654812
Protease	Gamma-Secretase	0.735093	0.568424
Protease	Cathepsin S	0.86755	0.439025
Protease	Dipeptidyl Peptidase IV	0.925892	0.350131
Protease	Beta-Secretase 1	0.958189	0.214969
Protease	Matrix Metalloproteinase-2	0.965923	0.501818
Protease	Renin	0.973206	0.409774
Protease	Matrix Metalloproteinase 13	0.980175	0.400279
Protease	Thrombin	1.01231	0.361709
Protease	Trypsin I	1.03733	0.622753
Protease	Human Immunodeficiency Virus Type 1 Protease	1.19677	0.375027
Protease	Leukocyte Elastase	1.24264	0.703556
Transporter	GABA Transporter 1	0.532831	0.68282
Transporter	Potassium-Transporting ATPase	0.551369	0.555607
Transporter	Norepinephrine Transporter	0.809101	0.369478
Transporter	Dopamine Transporter	0.929396	0.356622
Transporter	Serotonin Transporter	1.04327	0.252044
Others	Apoptosis Regulator BCL-2	0.640531	0.755704
Others	Bromodomain-Containing Protein 4	0.686667	0.461269
Others	Histone Deacetylase 1	0.739618	0.329351
Others	Histone Deacetylase 6	0.860364	0.345978
Others	P53-Binding Protein MDM-2	0.869291	0.520405