CHAPTER 1

INTRODUCTION

This chapter includes the following subtopics, namely: (1) Rationale; (2) Theoretical Framework; (3) Conceptual Framework/Paradigm; (4) Statement of the problem; (5) Hypothesis (Optional); (6) Assumption (Optional); (7) Scope and Delimitation; and (8) Significance of the study.

1.1 Rationale

Drug target interaction (DTI) refers to an interaction between drug and protein, or known as a target or receptor, in the human body [1]. Understanding these interactions is essential in the drug development process [2]. According to Kim et al. [3], there are more than a million drug compounds that have the potential to become new or repurposed drugs. Meanwhile, the success rate in drug development from phase I clinical trials to therapeutic licensure was relatively low, at only 6.2% of 21,143 compounds [4]. Therefore, accurate prediction of the DTI is an important part of drug development to find candidate compounds at an early stage [5].

One approach to predict the DTI is through conventional approaches conducted in the laboratory. Unfortunately, identifying drug and receptor interactions with such approaches is full of challenges, such as being time-consuming and cost-expensive [6]. One of the alternative methods for predicting drug and receptor interactions is by extracting sequence information from the drug and receptor and then using the information to recognize the relationship between the sequence and interactions of the drug and receptor. Recent studies highlight deep learning as a potential, cost-effective, and time-efficient method for modeling drug-receptor interactions, demonstrating significant success in addressing the complex, non-linear tasks frequently found in biological and chemical processes [7, 8].

Several studies have been performed to predict sequence-based drug and receptor interactions with deep learning by defining the task as a binary classification problem [9–11]. However, this approach has several limitations, such as the lack of experimental valid positive data and the abundance of unvalidated negative data. This approach defines that the drug and receptor have no interaction if the interaction data between them does not exist. However, the lack of interaction data does not guarantee that there is no interaction between drug and receptor. As a result, negative data can be much more numerous, creating a significant class imbalance [12]. Similarly, when using benchmark datasets such as Davis and KIBA [13], [14], researchers defined affinity values above a threshold as interacting pairs [15]. This led to a significant volume of negative data and resulting an imbalanced class distribution [9]. This condition has a negative impact on the performance of the prediction model. Furthermore, in reality, the affinity value that represents the strength of the drug-receptor interaction is known as continuous data [16]. Therefore, to address these issues, interactions between drugs and receptors should be defined as a regression problem. Hence, the prediction is conducted on the affinity value of the drug and receptor interaction, which is commonly called Drug Target Affinity (DTA).

Several studies have been performed on sequence-based DTA prediction using deep learning. In 2018, Ozturk et al. proposed DeepDTA, which uses CNN to model the SMILES representation of the drug and the amino acid sequence of the receptor. The results were evaluated using the concordance index (CI), which measures how well the model can rank relevant interaction pairs. The DeepDTA model showed good evaluation results with CI values of 87% and 86% for the Davis and KIBA datasets, respectively [13]. Then, in 2019, Ozturk et al. also proposed WideDTA, a model similar to DeepDTA, except they added features besides SMILES and amino acid sequences, such as ligand max common substructure and protein motifs and domains. WideDTA was able to increase the CI value of DeepDTA by one percent in both datasets [17]. In 2022, Ghimire et al. modified the CNN structure by inserting self-attention and produced a CI score of 89% in both the Davis and KIBA datasets [18]. Another study was conducted by D'Souza et al. by building a DTA sequence-based prediction model with CNN by adding prior transformers to the drug representation, which resulted in a CI score of 86% [19]. Similar studies have been conducted by modifying the CNN architecture and adding features other than sequences, such as fingerprints, to improve the performance of DTA prediction models. In 2022, Chen et al. proposed MultiscaleDTA by utilizing multi-scale CNN and self-attention at each CNN layer. This model achieved a CI value of 89% on the Davis and KIBA datasets [20]. Meanwhile, in 2023, Zhu et al. proposed FingerDTA which enriches the representation of drugs and receptors by incorporating molecular fingerprints in their CNN model. This approach also performed well, with a CI value of 89% on the same dataset [21].

Apart from the representation of drug and receptor, the inter-interaction or mutual interaction between drug and receptor is also important in predicting DTA. The aforementioned studies combine both representations only with a simple concatenation. This method overlooks the aspects of mutual interaction between drug and receptor representations [22]. In 2020, Abbasi et al. proposed DeepCDA, which uses CNN-LSTM as a representation method and a two-sided attention mechanism to achieve mutual interaction. This method resulted in competitive performance, with a CI value of 89% and an r_m^2 value of 64% [23]. In 2023, Zhao et al. conducted a similar study by proposing a two-sided attention mechanism to model mutual interactions, namely AttentionDTA, achieving an r2 value of 74% [14]. In 2021, another study by Zeng et al. took a different approach in using attention for mutual interactions, designating drug representations as queries and receptor representations as keys and values and resulting in a CI of 89% [22]. In 2021, Mahdaddi et al. combined SMILES and amino acid sequences before processing them with a CNN-AbiLSTM model, resulting in a CI of 89% and an r_m^2 value of 66% [12].

The studies mentioned above used CNN or CNN-LSTM to represent drugs or receptors to predict DTA accurately. Unfortunately, the results still have not been effective due to the inability of the models to effectively capture specific patterns in drug and receptor sequences with limited data. This is mainly due to the challenge of investigating the interaction's complexity and the high-dimensional nature of the data, which leads to room for improvement. Hence, exploring other methods for drug and receptor representation is become necessary. Pre-trained language models, proven powerful in natural language processing (NLP), offer an alternative method for representing drugs and receptors [24]. Pre-trained models such as ChemBERTa-2 and ESM-2 can represent drug molecules on SMILES sequences and proteins on amino acids, respectively [25, 26]. Pre-trained models are useful because they allow the model to use the transferable information encoded in the weights that have been pre-trained with a large amount of data. Pre-trained models can represent sequences effectively, but they only focus on describing interactions within a sequence and ignore interactions between two different sequences, in this case, drug and receptor sequences. Understanding the mutual relationship between drug and receptor is also crucial in DTA [27]. Therefore, a multi-head attention mechanism can be used to model the mutual interaction. According to the literature, the attention mechanism can observe the relationship between drugs and receptors simultaneously, thus enabling a more comprehensive understanding of the input data.

This study aims to enhance the performance of the DTA prediction model by implementing two pre-trained language models to obtain drug and receptor representations, namely ChemBERTa-2 and ESM-2. Both models were considered because of their suitability for the type of input data sequences, specifically SMILES and amino acid sequences [25, 26]. Moreover, ChemBERTa-2 has shown improved performance in capturing nuanced molecular information from a big collection of SMILES sequences [25]. Also, ESM-2 is able to understand the complex interactions between proteins from a bunch of amino acid sequences [26]. Furthermore, this study used a gated two-sided multi-head cross-attention mechanism (GMHA) to model the mutual interactions between drugs and receptors, allowing the model to simultaneously include within-sequence interactions and interactions between two different sequence types. In this study, GMHA modifies the multi-head attention mechanism in which we added dynamic scaling and a gate process. We added a learnable parameter to allow more flexibility in adjusting the level of scaling during training [28]. Also, we added a gate process inspired by the idea of the output gate in the study [29] after the linear and dropout layers to control how much proportion of the attention output and original input is included in the final result of GMHA. Finally, this study uses four fully connected layers to predict the affinity values of drug and receptor sequences using MSE as a loss function.

1.2 Theoretical Framework

The theoretical foundation of this study integrates several theories to address the problem of drug target affinity prediction (DTA) by using pre-trained language models and a gated multi-head attention mechanism. The theoretical framework includes the following:

1. Drug Target Interaction (DTI)

   Drug Target Interaction (DTI) is an interaction between a drug and a target, which can also be referred to as a protein or receptor. A drug and receptor pair is said to interact when the drug binds to a protein or receptor in the human body. The drug binds to the receptor in order to modulate the activity or function of the receptor. This interaction can result in desired pharmacological effects, such as disease treatment or symptom relief. In related studies, DTI is often formulated as a classification task, where drug and receptor pairs are categorized into two classes: interacting or non-interacting.

2. Drug Target Affinity (DTA)

   Drug Target Affinity or DTA refers to the level of interaction strength between a drug and its target, which is usually expressed as a continuous value. Therefore, DTA is generally formulated as a regression task. This affinity value reflects how strongly a drug can bind to its target receptor or protein, making it an important indicator in drug development. A low K_d value indicates that the interaction between the drug and the target receptor is stronger. Therefore, higher affinity values in pK_d and $KIBAscore$ indicate that the drug has a better ability to bind to the receptor.

3. Representation of Drug and Receptor Sequences

   Drug data can be represented using SMILES (Simplified Molecular Input Line Entry System) sequences, which is a text format that uniquely describes chemical structures. Meanwhile, protein or receptor data is represented by amino acid sequences, which represent the linear arrangement of amino acid residues that compose the receptor. This representation allows drug and receptor data to be used in sequence-based deep learning models to understand the complex relationship between the pair.

4. Transfer Learning and Pre-trained Language Model

   Transfer learning is an approach where a model that has been trained on one task can be used to solve a new task with a smaller amount of data. In the case of this study, pre-trained language models such as ChemBERTa-2 for representing drugs and ESM-2 for receptors were used. Both models are able to capture the complex features of SMILES and amino acid sequences based on previous training on large data sets, thus improving the model's ability to understand the characteristics of drug and receptor molecules.

5. Gated Multi-head Attention (GMHA)

   Gated Multi-Head Attention (GMHA) is an improvement of the multi-head attention mechanism that aims to improve the model's focus on important features. The attention mechanism allows the model to highlight relevant parts of the input data. In this study, GMHA is used to capture the mutual interaction between drug and receptor representations. Unlike the regular MHA, GMHA adds a gate mechanism to regulate the contribution between the attention result and the original input, thus helping the model prioritize key features in drug and receptor interactions.

1.3 Conceptual Framework/Paradigm

This study aims to predict the affinity between drug and target receptor (DTA) through a pre-trained language model representation approach and gated multi-head attention (GMHA) mechanism. The conceptual framework of this study includes several variables and processes, as shown in Figure 1.1.

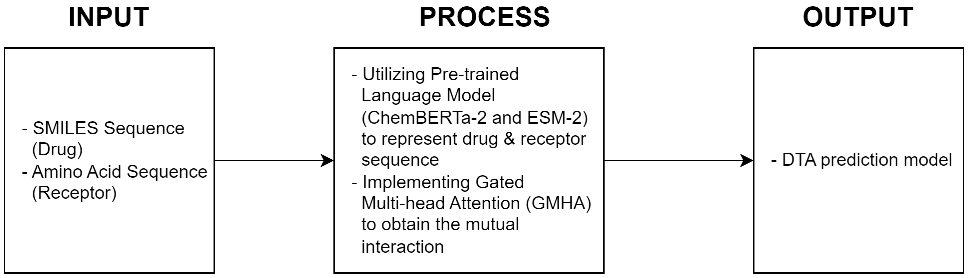
INPUT
PROCESS
OUTPUT
- SMILES Sequence (Drug)
- Amino Acid Sequence (Receptor)
- Utilizing Pre-trained Language Model (ChemBERTa-2 and ESM-2) to represent drug & receptor sequence
- Implementing Gated Multi-head Attention (GMHA) to obtain the mutual interaction
- DTA prediction model

Figure 1.1: The conceptual framework of this study

1. Variable Input

   The input in this study consists of two types of data, namely drugs represented by SMILES sequence and receptors represented by amino acid sequence. Both inputs are sequence data in string format.

2. Process

   The process in this study involves two main sequential steps. First, pre-trained language models, specifically ChemBERTa-2 and ESM-2, are utilized to extract important molecular features from SMILES and amino acid sequences. Second, gated multi-head attention (GMHA) is implemented to capture mutual interactions between drugs and receptors that ChemBERTa-2 and ESM-2 did not previously consider. The GMHA process prioritizes important features in such mutual interactions.

3. Variable Output

The output of this study is a DTA prediction model that is able to predict the binding affinity between the drug and the receptor in the form of a continuous value such as pK_d for Davis or $KIBAscore$ for KIBA. This value reveals the strong interaction between the drug and its target receptor.

1.4 Statement of the Problem

This study focuses on the challenge of predicting the affinity between drug and protein target (DTA), which is important in the drug discovery process. Current deep learning approaches still have limitations in performing accurate DTA predictions due to the lack of rich feature representation and the ability to model complex mutual interactions. This study attempts to answer three main questions:

1. How do the parameters of CEMDTA (proposed method) influence the performance of drug-target affinity prediction?
2. How does utilizing pre-trained language models and gated multi-head attention mechanisms influence the performance of drug-target affinity prediction?
3. How does the proposed method perform compared to baseline and existing benchmark models in DTA prediction?

This study should produce a more effective DTA prediction model by answering these three questions.

1.5 Objectives and Hypothesis

1.5.1 Objectives

There are three objectives based on research questions with a focus on improving the predictive ability of the prediction model as measured by MSE, CI, and r_m^2 across benchmark datasets, Davis and KIBA, namely:

1. To explore how CEMDTA (proposed method) parameters influence the performance of drug-target affinity prediction
2. To examine the impact of pre-trained language models and GMHA mechanisms on drug-target affinity prediction performance
3. To analyze the performance of the proposed method compared to the baseline and existing benchmark models

1.5.2 Hypothesis

The hypothesis to be tested is that the integration of ChemBERTa-2, ESM-2, and gated multi-head attention will result in a significant enhancement of the model's predictive ability in predicting DTA, as demonstrated by improved CI and r_m^2 on the specified benchmark datasets, Davis and KIBA.

1.6 Assumption

The implementation of pre-trained language models (ChemBERTa-2 and ESM-2) and gated multi-head attention is based on the following assumptions:

1. We assume that drug SMILES sequences and receptor amino acid sequences provide enough information to model drug target affinity accurately.
2. The datasets used in this study, Davis and KIBA, are assumed to be representative of real-world drug target affinity scenarios, providing a valid basis for evaluating the proposed model.

1.7 Scope and Delimitation

The main focus of this study is the drug target affinity (DTA) prediction using sequence-based deep learning models. Specifically, it uses SMILES sequences for drugs and amino acid sequences for receptors. The model utilizes pre-trained language models, namely ChemBERTa-2 and ESM-2, for extracting the important features and a gated multi-head attention (GMHA) mechanism to capture the mutual interactions between drug and receptor. The datasets used are limited to Davis and KIBA, the benchmark datasets in DTA prediction studies. These datasets allow for a fair comparison of model performance but may only cover some variations of drug target affinity in the real world. This study is also limited to sequence-based approaches without integrating data from other modalities, such as property or graph data.

1.8 Significance of the Study

The main contribution of this study is that we propose a novel architecture that leverages two pre-trained language models, ChemBERTa-2 and ESM-2, to obtain drug and receptor representations. These pre-trained language models capture important features from drug and receptor sequences. Then, we utilized the gated multi-head attention mechanism, specifically gated two-sided multi-head cross-attention, to simultaneously capture the mutual interactions between drugs and receptors. We introduced modifications such as dynamic scaling and gating to enhance this mechanism. Dynamic scaling involves learnable parameters that provide flexibility in adjusting the scaling factor during training. Meanwhile, the gating mechanism controls the proportion of attention output and incorporates the original input into the final result. This novel architecture aims to enhance the performance of deep learning models for drug target affinity (DTA) prediction.