DRUG BIOACTIVITY PREDICTION FOR ALZHEIMER'S DISEASE USING INTELLIGENT TECHNIQUES AND INTERACTION NETWORKS

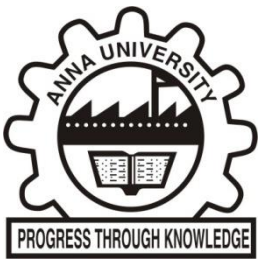
A SYNOPSIS

Submitted by

HEMKIRAN S

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY



FACULTY OF INFORMATION AND COMMUNICATION ENGINEERING

ANNA UNIVERSITY

CHENNAI 600 025

OCTOBER 2022

1. INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disorder, and is a leading cause of dementia cases in the world today. It is estimated that in excess of 100 million people will be affected by this dreadful disease by 2050 (Sadigh-Eteghad et al. 2015). AD adversely affects the functioning of the brain leading to impediments such as loss of memory, depression, swings in behavior, arduousness in performing daily activities, etc. Detecting AD at an early stage can retard the adverse effects that may manifest later in patients (Khan & Zubair 2019).

Several studies attribute the occurrence of AD to two hypotheses. The first, amyloid hypothesis propounds that the incidence of AD is due to the build-up of β-amyloid (Aβ) proteins in the form of plaques that accumulate in the neurons of the human brain. The second, cholinergic hypothesis proposes that the dysfunction of acetylcholine which transmits information between neurons, causes rapid degeneration of cognitive abilities, leading to AD. Although several therapies have mitigated the symptoms, no clinical methods have been identified thus far, to effectively treat AD (Long & Holtzman 2019). Therefore, identifying specific drugs to cure AD is paramount.

2. MOTIVATION

The process of Drug Discovery (DD) involves detecting new compounds to cure diseases. DD entails identifying new medications having high bioactivity against intended therapeutic targets, by conducting in-vitro and in-vivo experiments. However, this process of discovery is protracted and expensive. Consequently, in-silico methods that employ computational approaches to detect novel compounds, are essential. Virtual screening is the foremost step to identify lead molecules during in-silico drug design. This computational method can aid in rapidly determining possible lead compounds from public pharmacological repositories such as ChEMBL and ZINC

databases. Consequently, this dissertation investigates the use of supervised and unsupervised learning algorithms to precisely detect prospective lead compounds from the two databases in order to treat AD.

3. OBJECTIVES

The salient objectives of this dissertation are:

- To utilize Machine Learning (ML) / Deep Learning (DL) approaches and construct models to predict drug bioactivity.
- To employ bioactivity networks and visualize the interactions in order to identify bioactive compounds.

ML/DL techniques have been deployed because these supervised methods can efficiently detect bioactive drugs and predict their effectiveness to cure AD at an early stage (Chitradevi & Prabha 2020). Further, bioactivity networks are constructed employing unsupervised clustering techniques to envisage the interactions among drug compounds. Both these objectives are attained utilizing two publicly available biological databases namely, ChEMBL and ZINC. The research framework employed in this dissertation to realize the objectives, is elucidated in the forthcoming section.

4. FRAMEWORK OF THE RESEARCH

The overall framework of the research is depicted in Figure 4.1. Initially, the requisite dataset of drug compounds pertaining to the Aβ and cholinergic hypothesis is obtained from the ChEMBL and ZINC databases respectively. The next step entails transforming, curating and refining the data to generate a precise dataset with no missing values or duplicates. The Simplified Molecular Input Line Entry System (SMILES) notation of the drugs in the dataset is utilized to ascertain the bioactivity, by filtering the drugs which adhere to the Lipinski's Rule of Five (LRo5).

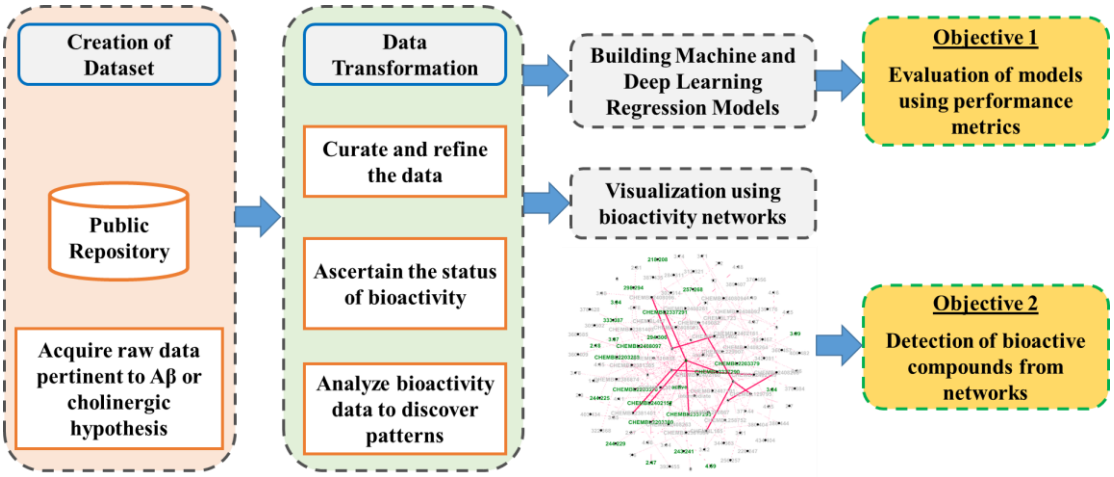


Figure 1. Research Framework

Thereafter, to attain the first objective, supervised learning approaches are utilized to build ML regression models to predict drug bioactivity and their performance is compared with that of DL models. The second objective namely, the identification of bioactive compounds is accomplished by employing unsupervised clustering techniques and constructing bioactivity networks of compounds. The research framework is further delineated in section 5 hereunder.

5. PROPOSED METHODOLOGIES

In this section the methodologies adopted to attain the two objectives specified in the research framework are expounded. This synopsis exemplifies the procedure of accomplishing the objectives utilizing the ChEMBL database. The analogous process for ZINC database is elucidated in the dissertation.

5.1 Model Building and Evaluation (MBE)

The first objective of this dissertation namely MBE is to predict drug bioactivity using supervised learning techniques. Typically, ML is utilized in scenarios entailing efficient handling and interpretation of data, with minimal necessity for programming. This enhances the learning abilities of computers. Consequently, due to the emergence of heterogeneous datasets, ML is