UNIVERSITY OF GHANA-LEGON

DCIT 400 – FIRST CUT PROPOSAL

BY

ADDO STEPHEN TETTEH

10919773

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF A

B.Sc. (Hons) DEGREE IN INFORMATION TECHNOLOGY.

December 2023

A PREDICTIVE MACHINE LEARNING MODEL FOR BREAST CANCER DRUG-TARGET INTERACTIONS

1. Introduction:

Identifying potential drug candidates is ineffective and slow, and the chance of failing clinical trials after discovery is relatively high. This problem increases the time and resources used by researchers to obtain drugs with clinical efficacy for disorders like breast cancer. A machine learning (ML)-based web application is a proposed solution that will empower researchers and clinicians with a rapid and cost-effective method for prioritizing promising drug candidates and informing treatment decisions.

2. Aim and Specific Objectives:

Aim: To develop a machine learning model for predicting drug-target interactions (DTIs) relevant to breast cancer.

Specific Objectives:

1. To collect and curate high-quality datasets on breast cancer proteins.

2. To develop and train one ML model (random forests) for PIC50 bioactivity for DTI prediction using curated data.

3. Develop a user-friendly web application that allows researchers and clinicians to easily access and utilize the model for predicting breast cancer drug-target interactions.

4. Evaluate and validate the model's performance using rigorous cross-validation techniques and independent validation datasets.

5. To evaluate the web application's usability and impact through user feedback.

3. Methodology:

The project will employ a multi-pronged approach:

- Data acquisition: Public databases such as CHEMBL will be utilized to obtain information and known DTIs for breast cancer.

- Data preprocessing: Missing values will be imputed, features will be engineered, and data will be normalized for ML model training.
- Model development: Two ML algorithms will be explored and compared based on their performance on benchmark datasets.
- Web application development: A user-friendly interface will be built using appropriate web development frameworks and tools.

4. Expected Outcomes:

At the end of this project, I expect to have a high-performing predictive ML model for DTIs relevant to breast cancer to aid researchers and clinicians in improving personalized treatment strategies via our user-friendly web application.