

CHAPTER – 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 Introduction

Personalized medicine is gaining prominence in cancer treatment due to the diverse genetic makeup of cancer cells, which leads to varied responses among patients undergoing similar therapies. This project, titled “Leveraging ANN for Targeted Drug Sensitivity Prediction on GDSC Data”, aims to create a model that leverages genomic data to predict how different cancer cell lines respond to specific drugs. The complexity of cancer biology and the uniqueness of each patient’s genetic profile necessitate computational tools that can support precision oncology by identifying the most effective drug treatments. This predictive model, based on the ANN architecture, seeks to aid oncologists and researchers by providing tailored predictions of drug efficacy based on specific cancer genomics, thus advancing the approach toward personalized therapies.

Our project uses the Genomics of Drug Sensitivity in Cancer (GDSC) dataset, which contains essential genomic features and drug response data across over 1,000 cancer cell lines. Focusing on IC50 prediction, a critical metric for assessing drug potency, the model leverages features such as gene mutations, expression profiles, and cancer-type descriptors to capture relevant patterns in drug sensitivity. A user-friendly web-based interface developed with Flask allows researchers to input data for IC50 predictions seamlessly. By simplifying the interaction between the model and users, this web application supports data input through dropdown menus and displays prediction outputs in an accessible, clinician-oriented manner, aiming to streamline and support decision-making in cancer research and clinical practice.

1.2 Problem Statement

Cancer treatment efficacy is significantly impacted by the genetic diversity of cancer cells, causing variable responses to the same drug among patients. Traditional treatment approaches often do not account for these individual differences, leading to inconsistent therapeutic outcomes. Furthermore, the lack of reliable predictive models for drug response limits the scope of precision medicine. This project addresses the following key issues.

- **Limited personalized treatment options:** Conventional treatments do not consider individual genetic profiles.
- **Challenges in predicting drug efficacy:** The relationship between genomic features and drug response is complex and not fully understood.
- **High failure rates in therapy selection:** The absence of predictive accuracy can lead to ineffective treatments, higher costs, and increased side effects.

To tackle these challenges, our project proposes an ANN-based model trained on genomic data to predict drug sensitivity, helping oncologists make better-informed treatment decisions. By analyzing genomic features relevant to drug response, the model aims to provide more accurate predictions that align with the unique profiles of cancer patients.

1.3 Objectives

- Merge data from the GDSC, gene expression, mutation profiles, and CNAs to enhance the accuracy and relevance of drug response predictions.
- Build an ANN model to predict IC50 values based on genomic data from multiple datasets.
- Develop a Flask-based interface to facilitate easy data input.
- Enable users to visualize IC50 predictions along with a comparative graph for five selected drugs.

1.4 Scope

- **Data Analysis:** Analyze genomic features from the GDSC dataset to identify correlations with drug response, such as gene expression and mutations.
- **Model Development:** Implement an ANN model to capture complex relationships between genomic features and drug efficacy.
- **Performance Evaluation:** Assess model performance using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE) and R-squared (R^2).
- **Web Interface Development:** Design a Flask-based web interface to facilitate user interaction with the model and simplify data entry.
- **Real-World Application:** Develop the interface with real-time prediction capabilities, suitable for clinical and research use cases.

1.5 Organization of the Report

This report is organized as follows:

- **Chapter 1:** Introduction, which includes the background, problem statement, objectives, scope, and structure of the report.
- **Chapter 2:** Literature Survey, providing an overview of existing systems and approaches in drug response prediction and discussing their limitations.
- **Chapter 3:** System Requirements Specification, defining the functional and non-functional requirements of the project.
- **Chapter 4:** Gantt Chart, displaying the project timeline and key milestones.
- **Chapter 5:** System Design, describing the architecture, data flow diagrams, and use case diagrams of the proposed solution.
- **Chapter 6:** Implementation, covering data preprocessing, model architecture, and the Flask application development process.
- **Chapter 7:** System Testing, outlining the testing methodologies, test cases, and validation techniques used to evaluate the model.
- **Chapter 8:** Results and Snapshots, presenting model performance metrics and screenshots of the Flask interface.
- **Chapter 9:** Conclusion and Future Work, summarizing project outcomes and suggesting future directions for research.
- **References:** Listing the sources consulted throughout the development of the project.