##### A Project report on

**DRUG RESPONSE PREDICTION USING XGBOOST**

###### A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

**Bachelor of Technology**

**in**

**Computer Science and Engineering**

Submitted by

B.PRAVEEN KUMAR

(20H51A05D8)

B.NARESH

(20H51A0532)

V.VAMSHI

( 20H51A05M6)

Under the esteemed guidance of

Dr.P. Senthil

(Assistant Professor)



**Department of Computer Science and Engineering**

**CMR COLLEGE OF ENGINEERING & TECHNOLOGY**

(UGC Autonomous)

\*Approved by AICTE \*Affiliated to JNTUH \*NAAC Accredited with A+ Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

#### 2020- 2024

**CMR COLLEGE OF ENGINEERING & TECHNOLOGY**

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



#### CERTIFICATE

This is to certify that the Major Project Phase I report entitled **"GCN BASED DRUG RESPONSE PREDICTION"** being submitted by B.PraveenKumar (20H51A05D8), B. Naresh (20H51A0532), V. Vamshi (20H51A05M6), in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

###### The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

**Dr.P. Senthil Dr. Siva Skandha Sanagala**

**Assistant Professor Associate Professor and HOD**

**Dept. of CSE Dept. of CSE**

#### ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Dr.P. Senthil, Assistant professor**, Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

B. Praveen kumar - 20H51A05D8

B. Naresh - 20H51A0532

V.Vamshi - 20H51A05M6

GCN BASED DRUG RESPONSE PREDICTION

**TABLE OF CONTENTS**

**CHAPTER**

**NO. TITLE PAGE NO.**

LIST OF FIGURES 11,13

ABSTRACT 4

**1** **INTRODUCTION** 5-6

1.1 Problem Statement 6

1.2 Research Objective 6

**2** **BACKGROUND WORK** 7-13

2.1. Existing system 8

2.1.1.Introduction 8

2.1.2.Merits,Demerits and Challenges 8

2.2. Proposed system

2.2.1.Introduction 9

2.2.2.Merits,Demerits and Challenges 9

2.2.3.Implementation of Existing Method 10-13

**3 RESULTS AND DISCUSSION 14-15** 3.1 Performance metrics 15

**4** **CONCLUSION 16-17**

4.1 Conclusion 17

**REFERENCES** 18-19

**GitHub Link 20**

CMRCET B. Tech (CSE) Page No 1

GCN BASED DRUG RESPONSE PREDICTION

**List of Figures**

**FIGURE**

**NO. TITLE PAGE NO.**

2.1 Architecture Diagram 11

2.2 Sequence Diagram 13

CMRCET B. Tech (CSE) Page No 2

GCN BASED DRUG REPONSE PREDICTION

# **ABSTRACT**

* Drug response prediction is an important problem in computational personalized medicine. Many machine-learning-based methods, especially deep learning-based ones, have been proposed for this task. However, these methods often represent the drugs as strings, which are not a natural way to depict molecules. Also, interpretation (e.g., what are the mutation or copy number aberration contributing to the drug response) has not been considered thoroughly.

CMRCET B. Tech (CSE) Page No 4

GCN BASED DRUG REPONSE PREDICTION

# **CHAPTER 1**

**INTRODUCTION**

CMRCET B. Tech (CSE) Page No 5

GCN BASED DRUG RESPONSE PREDICTION

**CHAPTER 1**

**INTRODUCTION**

**1.1.Problem Statement**

* Using the right drug in the right dose at the right time is a goal in personalized medicine. Thus, estimating how each patient responds to a drug based on their biological characteristics (e.g., omics data) is important in biomedical research. However, patients’ drug response data is very limited and not well-structured. Indeed, there have been only a few studies on drug response for cancer patients gathered in TCGA [1]. This has formed a barrier to large scale research on this topic.

**1.2.Research Objective**

* **Early Detection of Malware Activities:** The primary research objective is to enable the early detection of malware activities in the context of cyber attacks.
* **Anomalous Synchronization:** The focus of the research is on identifying and understanding anomalous synchronization patterns within spatiotemporal dark net traffic data.
* **Machine Learning Methods:** The study aims to develop and utilize three independent machine learning methods to automatically estimate and detect these anomalous patterns in real-time.
* **Comprehensive Framework:** The research aims to create a comprehensive framework that combines these methods, leveraging their complementary strengths,to improve the accuracy of detection.
* **Achieving 100% Recall Rate:** The ultimate research goal is to achieve a 100% recall rate, indicating a highly effective approach

CMRCET B. Tech (CSE) Page No 6

GCN BASED DRUG RESPONSE PREDICTION

**CHAPTER 2**

**BACKGROUND WORK**

CMRCET B. Tech (CSE) Page No 7

GCN BASED DRUG RESPONSE PREDICTION

**CHAPTER 2**

**BACKGROUND WORK**

**2.1.EXISTING SYSTEM :**

**2.1.1.INTRODUCTION :**

* The development of new drugs is costly, time consuming and often accompanied with safety issues. Drug repurposing can avoid the expensive and lengthy process of drug development by finding new uses for already approved drugs. In order to repurpose drugs effectively, it is useful to know which proteins are targeted by which drugs. Computational models that estimate the interaction strength of new drug-target pairs have the potential to expedite drug repurposing. Several models have been proposed for this task. However, these models represent the drugs as strings, which is not a natural way to represent molecules.
* An existing system defines a new model called GraphDTA that represents drugs as graphs and uses graph neural networks to predict drug-target affinity. We show that graph neural networks not only predict drug-target affinity better than non-deep learning models, but also outperform competing deep learning methods. Our results confirm that deep learning models are appropriate for drug-target binding affinity prediction, and that representing drugs as graphs can lead to further improvements.

**2.1.2. Disadvantages :**

* The system is not implemented Prediction of unknown drug-cell line response of the datasets.
* The system is not implemented graph Convolutional network for drug response prediction (graphdrp).

CMRCET B. Tech (CSE) Page No 8

GCN BASED DRUG RESPONSE PREDICTION

2.2.Proposed System :

**2.2.1.Introduction :**

* In this study, the system proposes GraphDRP (Graph convolutional network for drug response prediction), a new neural network architecture capable of modeling drugs as molecular graphs to predict drug response on cell-line. We compared our method with the state-of-the-art, tCNNS [33], where drug molecules were represented as SMILES [39] strings.
* Experimental results indicate that our method achieves better performance in terms of root mean square error (RMSE) and Pearson correlation coefficient for all experiments. Also, by visualizing the resulting networks through saliency maps, we can discover the most significant genomic aberrations for the prediction of the response value. This suggests a novel way to interpret the result of deep learning models for drug response prediction.

**Advantages**

* The system is implemented GRAPH CONVOLUTIONAL NETWORK FOR DRUG RESPONSE PREDICTION (GRAPHDRP).
* The proposed system aims at predicting missing drug-cell line response. The best pre-trained model in the mixed test experiment was used to predict missing pairs in GDSC dataset.

CMRCET B. Tech (CSE) Page No 9

GCN BASED DRUG RESPONSE PREDICTION

**SYSTEMREQUIREMES**

**H/W System Configuration:-**

* **Processor**  - Pentium –IV
* **RAM**  - 4 GB (min)
* **Hard Disk**  - 20 GB
* **Key Board**  - Standard Windows Keyboard
* **Mouse**  - Two or Three Button Mouse
* **Monitor**  - SVGA

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 7 Ultimate.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Back-End :** Django-ORM
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL (WAMP Server).

CMRCET B. Tech (CSE) Page No 10

GCN BASED DRUG RESPONSE PREDICTION

**Architecture Diagram**

Service Provider

Login,

Browse and Train & Test Data Sets,

View Trained and Tested Accuracy in Bar Chart,

View Trained and Tested Accuracy Results,

View Prediction Of Drug Response,

View Drug Response Ratio,

Download Trained Data Sets,

View Drug Response Ratio Results,

View All Remote Users .

**Web Server**

Accepting all Information

Datasets Results Storage

Process all user queries

Accessing Data

**Store and retrievals**

**WEB Database**

REGISTER AND LOGIN,

PREDICT DRUG RESPONSE TYPE,

VIEW YOUR PROFILE.

Remote User

Tweet Server

Tweet Server

Tweet Server

FIG 2.1- Architecture diagram

CMRCET B. Tech (CSE) Page No 11

GCN BASED DRUG RESPONSE PREDICTION

**Modules**

**Service provider** :

* In this module, the Service Provider has to login by using valid user name and password After login successful he can do some operations such as Login, Browse and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Drug Response, View Drug Response Ratio, Download Trained Data Sets, View Drug Response Ratio Results, View All Remote Users.

**View and Authorize Users :**

* In this module, the admin can view the list of users who all registered. In this, the admin can view the user’s details such as, user name, email, address and admin authorizes the users.

**Remote User :**

* In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT DRUG RESPONSE TYPE, VIEW YOUR PROFILE.

CMRCET B. Tech (CSE) Page No 12

GCN BASED DRUG RESPONSE PREDICTION

* **Sequence Diagram**

Service Provider

Remote User

Web Server



REGISTER AND LOGIN

PREDICT DRUG RESPONSE TYPE

VIEW YOUR PROFILE

Login,

Browse and Train & Test Data Sets,

View Trained and Tested Accuracy in Bar Chart,

View Trained and Tested Accuracy Results,

View Prediction Of Drug Response,

View Drug Response Ratio,

Download Trained Data Sets,

View Drug Response Ratio Results,

View All Remote Users.

FIG 2.2

CMRCET B. Tech (CSE) Page No 13

GCN BASED DRUG RESPONSE PREDICTION

**CHAPTER 3**

**RESULTS AND DISCUSSION**

CMRCET B. Tech (CSE) Page No 14

GCN BASED DRUG RESPONSE PREDICTION

**CHAPTER 3**

**RESULTS AND DISCUSSION**

* In this study, we proposed a novel method for drug response prediction, called GraphDRP. In our model, drug molecules were presented as graphs instead of strings, celllines were encoded into one-hot vector format. Then graph convolutional layers were used to learn the features of compounds and 1D convolutional layers were used to learn cell-line representation.After that the combination of drug and cell-line representation was used to predict IC50 value.
* Four variants of graph neural networks including GCN, GAT, GIN and combination of GAT&GCN were used for learning features of drugs. We compared our method with state-of-the-art one, tCNNS [33], where drug molecules were represented as SMILES strings. Experimental results indicate that our method achieves better performance in terms of both root mean square error and Pearson correlation coefficient.
* The performance suggests that representing drugs in graphs is more suitable than in strings since it conserves the nature of chemical structures of drugs. Furthermore, the responses of missing drug-cell line pairs in GDSC dataset were predicted and analyzed. We figured out that Bortezomib and Epothilone B have the lowest IC50 values and we found the evidence showing that some types of cancer are sensitive to these drugs. Similarly, we also found that cancers are less sensitive to drugs having the highest IC50 values.
* It means that the model actually learns from data and has a potential to predict the response of new drug-cell line pairs. Also, through saliency maps, we discovered ten most important genomic aberrations of the three cell-lines having lowest IC50s to that drug and seek their contribution to the sensitivity of that drug. This technique suggests a novel way to interpret the result of deep learning model in drug response.

CMRCET B. Tech (CSE) Page No 15

GCN BASED DRUG RESPONSE PREDICTION

CHAPTER 4

**CONCLUSION**

CMRCET B. Tech (CSE) Page No 16

GCN BASED DRUG RESPONSE PREDICTION

**CHAPTER 4**

**CONCLUSION**

* In this study, we proposed a novel method for drug response prediction, called Graph DRP. In our model, drug molecules were presented as graphs instead of strings, cellines were encoded into one-hot vector format. Then graph convolutional layers were used to learn the features of compounds and 1D convolutional layers were used to learn cell-line representation. After that the combination of drug and cell-line representation was used to predict IC50 value.
* Four variants of graph neural networks including GCN, GAT, GIN and combination of GAT&GCN were used for learning features of drugs. We compared our method with state-of-the-art one, TCNNS [33], where drug molecules were represented as SMILES strings. Experimental results indicate that our method achieves better performance in terms of both root mean square error and Pearson correlation coefficient.

CMRCET B. Tech (CSE) Page No 17

GCN BASED DRUG RESPONSE PREDICTION

**REFERENCES**

CMRCET B. Tech (CSE) Page No 18

GCN BASED DRUG RESPONSE PREDICTION

## REFERENCES

[1] J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. M. Shaw, B. A.Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J. M. Stuart,C. G. A. R. Network et al., “The Cancer Genome Atlas Pan-Cancer analysis project,” Nature Genetics, vol. 45, no. 10, p. 1113, 2013.

[2] W. Yang, J. Soares, P. Greninger, E. J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D.Beare, J. A. Smith, I. R. Thompson et al.,“Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells,” Nucleic Acids Research, vol. 41, no. D1, pp. D955–D961, 2012.

[3] J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C.J.Wilson, J. Leh´ar, G. V. Kryukov, D. Sonkin et al.,“The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity,” Nature, vol. 483, no. 7391, pp. 603–607, 2012.

[4] R. H. Shoemaker, “The NCI60 human tumour cell line anticancer drug screen,” Nature Reviews Cancer, vol. 6, no. 10, pp. 813–823, 2006.

[5] F. Azuaje, “Computational models for predicting drug responses in cancer research,” Briefings in Bioinformatics, vol. 18, no. 5, pp. 820–829, 2017.

[6] J. Chen and L. Zhang, “A survey and systematic assessment of computational methods for drug response prediction,” Briefings in Bioinformatics, 01 2020, bbz164. [Online]. Available:https://doi.org/10.1093/bib/bbz164

[7] M. J. Garnett, E. J. Edelman, S. J. Heidorn, C. D. Greenman, A. Dastur, K. W. Lau, P.Greninger, I. R. Thompson, X. Luo, J. Soares et al., “Systematic identification of genomic markers of drug sensitivity in cancer cells,” Nature, vol. 483, no. 7391, pp. 570–575, 2012.

[8] I. S. Jang, E. C. Neto, J. Guinney, S. H. Friend, and A. A. Margolin, “Systematiassessment of analytical methods for drug sensitivity prediction from cancer cell line data,” in Biocomputing. World Scientific, 2014, pp. 63–74.

## 

CMRCET B. Tech (CSE) Page No 19

**GitHub Link**

1. https://github.com/BANOTH-NARESH/DRUG-RESPONSE-PREDICTION-USING-XGBOOST/