How to Cite:
Savitha, P. M., & Rani, M. P. (2022). A novel algorithm for predicting adverse drug-drug interaction. *International Journal of Health Sciences*, *6*(S4), 11564–11575. https://doi.org/10.53730/ijhs.v6nS4.11234

A novel algorithm for predicting adverse drug-drug interaction

P Margaret Savitha
P Margaret Savitha, Research scholar, Mother Teresa University, Kodaikanal, TamilNadu, India
Corresponding author email: margaretsavitha@gmail.com

Dr. M Pushpa Rani
Dr. M Pushpa Rani, Professor & Head, Mother Teresa University, Kodaikanal, TamilNadu, India
Email: drpushpa.mtwu@gmail.com

***Abstract*---**Molecules are the fundamental units of any chemical compound. A drug is a predefined composition of chemical compounds which is expected to produce a target biological behavior. Drug error is a term used to indicate a scenario where one drug increases or decreases the effect of another drug when both are present inside the human body. This may be either due to pharmacodynamics (effect or action of a drug) or pharmacokinetics (movement of a drug within the body) or both. The results are painful as a person can suffer from serious disability or health hazard. The problem is mainly due to human error while prescribing and lack of efficient systems that can aid in automatic diagnosis of interacting drugs. Most of the issues that a patient undergoes or becomes a victim due to polypharmacy can be avoided and prevented. What we lack is an intelligent system that can predict the adverse interaction of poly drugs and raise an alarm to the medical team. This can be a live saver. This paper attempts to propose a model based on machine learning that can predict adverse drug-drug interaction.

***Keywords*---**machine learning, algorithm, predictive, analytics, healthcare, adverse reactions, artificial intelligence, drug-drug interactions, DDI.

Introduction

Molecules are the fundamental units of any chemical compound. A drug is a predefined composition of chemical compounds which is expected to produce a target biological behavior. Drug error is a term used to indicate a scenario where

one drug increases or decreases the effect of another drug when both are present inside the human body. This may be either due to pharmacodynamics (effect or action of a drug) or pharmacokinetics (movement of a drug within the body) or both. The results are painful as a person can suffer from serious disability or health hazard. The problem is mainly due to human error while prescribing and lack of efficient systems that can aid in automatic diagnosis of interacting drugs. Most of the issues that a patient undergoes or becomes a victim due to polypharmacy can be avoided and prevented. What we lack is an intelligent system that can predict the adverse interaction of poly drugs and raise an alarm to the medical team. This can be a live saver.

Adverse drug-drug side effect

Adverse drug side effect is in fact an injury due to taking a drug. This could have happened due to a prolonged dosage of a single drug or combination of multiple drugs. The interest of this paper is on the latter where a harmful result is obtained because a person consumes more than one drug and they react with each other thereby causing deterioration in the health condition.

The adverse outcome could be any of the following

- Mortality
- Disability
- Life-threatening
- Abnormality
- Hospitalization

The adverse effect could be caused in a local region of the body or could have affect the functioning of an entire system in the body. Medical practioners term it as the risk due to polypharmacy. According to research nearly about 8 in 1000 adults aged above 65 are prone to adverse drug-drug side effects. As per Mckinsey the cost of about 35 million advese drug effects amounts to roughly $115 billion USD.

Categorization of drugs involved in Drug-Drug interactions

A drug that is part of a drug-drug interaction is generally categorized as sufferer or executioner depending upon the role that they actively play. A sufferer is one that is affected by drug-drug interaction and is also called as victim. An executioner is on that causes the drug-drug interactions and is also called as perpetrator. An executioner is further categorized as an inducer and a determent or inhibitor. An executioner which is of category inducer will increase the expression or function of metabolic activity while an executioner which is of category determent or inhibitor will caused the opposite action to reduce the expression or function of metabolic activity [4].

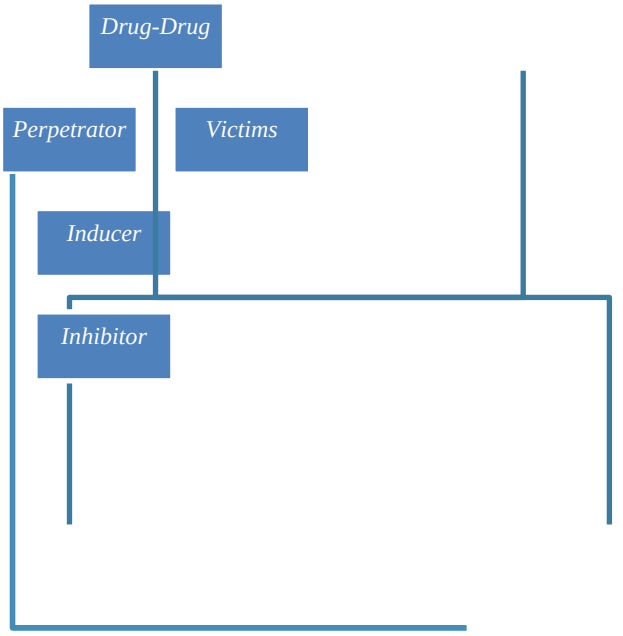

Fig.1

Literature survey

The most common public databases are Drugbank, Supertarget, Matador, Therapeutic target database, Sider and Kyoto encyclopedia of genes and genomes. The traditional methods for computing and predicting drug-target interactions are docking simulation, ligand-based methods and literature text-mining methods.

Disadvantages of the traditional computing methods

a. The docking simulation method needs the 3-D structure details of the target proteins. Unfortunately, the 3-D structure is not defined for all the proteins. This is a serious limitation.
b. Ligand is a molecule that binds a central metal atom to form a coordination complex[3]. The number of known ligands is very less due to which the applicability of the Ligand-based methods is very less.
c. Literature text-mining methods uses keywords to find the text to be mined. Keywords are elementary form of search and they are not very effective.

Summary of related work[1]

Author	Objective and approach
Monica Campillos, 2008	Predicted targets using side effect similarity and similarity networks
Edgar D Coelho, 2016	Predicted drug target interactions without considering type of target using

	Random forest, Support-Vector Machine and Logistic regression
Diego Galeano, 2016	Predicted targets based on drug chemical structure using Tanimoto similarity
Arvind Sinha, 2017	Predicted usability of the protein whether being a drug-target or a vaccine using Support-Vector Machine, Decision tree, Random forest and Naïve Bayes
Ming Hao, 2017	Predicted interactions between drugs and target using DNILMF
Ming Wen, 2017	Predicted new DTI without considering the type of targets using Deep learning
Hafez Manoochehri, 2018	Predicted drug-target interactions using DMF and KNN
Abdelrahman Saad, 2019	Predicted drug-target interactions using machine learning classification techniques Random Forest, Decision Tree and K-Nearest Neighbor
Abdekrahman Saad, Yasser Omar, Fahima Maghraby, 2019	Predicting drug interaction with Adenosine Receptors using Machine Learning and SMOTE techniques

Table. 1

Drug fingerprint

Drug fingerprint captures the molecule sequence information of a drug in its essence. It is basically a way of representing a drug. The drug fingerprint is a resultant of simulation of the molecule constituents of the drug, such as its atomic number and bonds between the atoms. A binary encoded code is generated which is assigned as the fingerprint of the drug. This can be used for classification and finding drug similarities during drug analysis.

Usage of drug fingerprints in prediction of drug-drug interactions

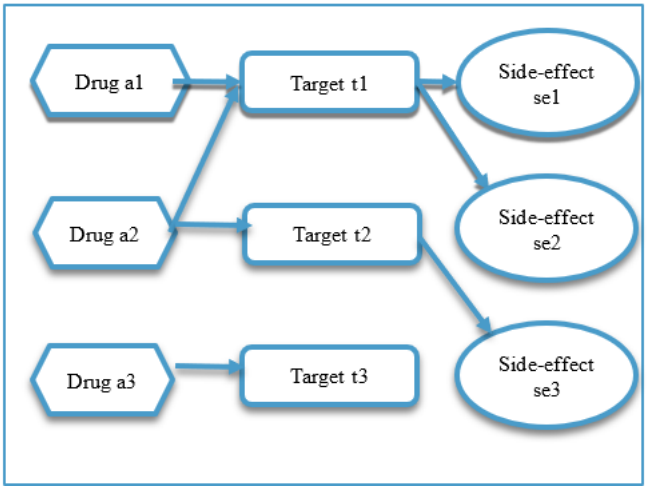

Fig 2

Drugs may interact with a single target or with multiple targets. The drug interaction can be represented as a triplet <Drug a1, Target t1, Side-effect se1>, <Drug a1, Target t1, Side-effect se2>. Drug fingerprint similarity analysis can be conducted to find and to also predict drugs that could interact with the same target, thereby aiding in the discovery of new application for a drug. Conversely, if the drug fingerprint is known then the possible drug interactions with a target and possible side-effects that it could cause can be predicted with a decent level of accuracy.

Position Specific Scoring Matrix (PSSM)

The position specific scoring matrix helps to retain the evolutionary information. The PSSM for Zinc (cd02249) is shown below:

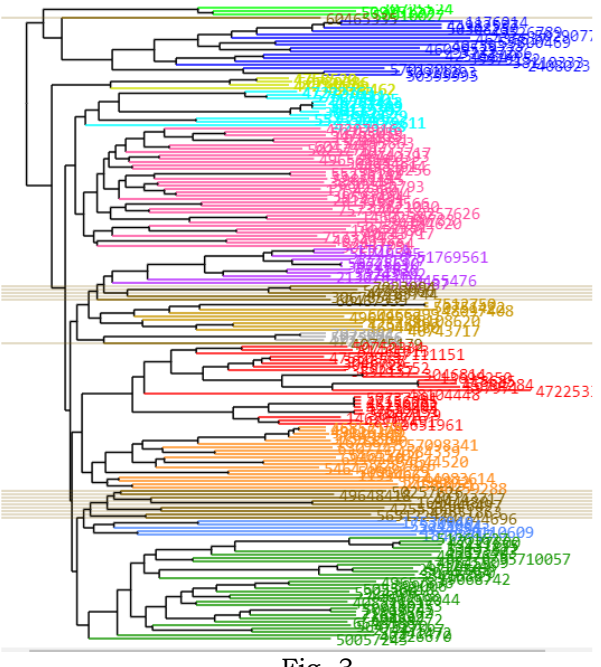

Fig. 3

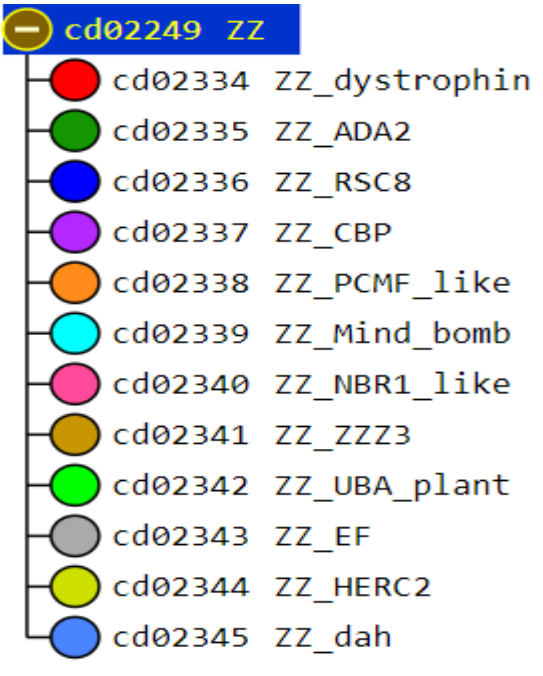
cd02249 ZZ
cd02334 ZZ_dystrophin
cd02335 ZZ_ADA2
cd02336 ZZ_RSC8
cd02337 ZZ_CBP
cd02338 ZZ_PCMF_like
cd02339 ZZ_Mind_bomb
cd02340 ZZ_NBR1_like
cd02341 ZZ_ZZZ3
cd02342 ZZ_UBA_plant
cd02343 ZZ_EF
cd02344 ZZ_HERC2
cd02345 ZZ_dah

Fig.4

P	C	Master	A	G	I	L	V	M	F	W	P	C	S	T	Y	N	Q	H	K	R	D	E
26	C	31 - C	-4	-6	-5	-5	-4	-5	-6	-6	-6	12	-4	-4	-6	-6	-7	-7	-7	-7	-7	-7
29	C	34 - C	-4	-6	-5	-5	-4	-5	-6	-6	-6	12	-4	-4	-6	-6	-7	-7	-7	-7	-7	-7
6	C	12 - C	-4	-6	-5	-5	-4	-5	-6	-6	-6	12	-4	-4	-6	-6	-7	-7	-7	-7	-7	-7
3	C	9 - C	-4	-6	-5	-5	-4	-5	-6	-6	-6	12	-4	-4	-6	-6	-7	-7	-7	-7	-7	-7
17	C	22 - C	-4	-6	-5	-5	-4	-5	-6	-6	-6	11	-1	-4	-6	1	-6	-6	-6	-7	-6	-7
20	C	25 - C	0	-6	-5	-5	-4	-5	-6	-6	-6	11	-4	-4	-6	-6	-6	-6	-1	-6	-7	-7
41	H	42 - H	-5	-1	-7	-6	-7	-5	-5	-6	-6	-7	-4	-5	-2	-3	2	11	-4	-4	-4	-1
37	H	40 - H	-5	-5	-7	-7	-7	-5	-5	-6	-6	-7	-4	-5	-2	4	-3	11	-4	-4	-4	-4
30	Y	35 - Y	-5	-7	-1	-1	-4	-4	6	-2	-7	-6	-5	-5	8	-6	-1	2	-6	-5	-7	-6
16	H	21 - H	-1	-5	-6	-6	-6	-5	-6	-6	-5	-6	-1	-1	-4	2	-3	8	4	2	-4	1

Fig. 5

Synthetic Minority Oversampling Technique (SMOTE) algorithm

SMOTE is a balancing technique used to rationalize the interacting drugs. This is usually used when the classification results are affected.

D syn = D i + (D Knn – D i) * r,

where
D syn = synthetic data
D i = minority samples
D Knn = sample of K-nearest neighbor from minority samples
r = random number between 0 and 1

The algorithm is as follows:

i. Find D i and D Knn
ii. Determine the difference between the feature vector K-nearest neighbor from the minority sample
iii. Multiply the result by a random number r
iv. Add the value to D i
v. Repeat steps i to iv for all new vectors

Measuring performance

For the sake of research, we have used only one parameter for measuring performance which is accuracy. The metric is calculated as follows:

Accuracy = [(True Positive + True Negative)] / [(True Positive + True Negative + False Positive + False Negative)], where

True Positive = Instance of an Adverse reaction
True Negative = Instance of normal intended reaction
False Positive = Predicted instance of an adverse reaction
False Negative = Predicted instance of a normal reaction

Results and Findings

The drug-drug interaction information was stored in a matrix format. A cell was marked as 1 if there is a history of adverse reaction between the chemical composition of the two named drugs. This information is obtained from the SIDER database

	Drug 1	Drug 2	Drug 3	Drug 4	Drug 5
Drug 1	0	0	0	1	1
Drug 2	0	0	1	0	0
Drug 3	0	1	0	0	0
Drug 4	1	0	0	0	1
Drug 5	1	0	0	1	0

Table 2

For testing purpose 5 drugs were selected from the Sider database. Their names have been masked as the intention of this study is to fine tune the algorithm that can be used for Machine Learning and not on the drug. In this case Drug1 exhibits adverse behavior when taken alongside Drug4 & 5, similarly Drug2 along with Drug3 and Drug4 with Drug5. The resultant matrix is Sparse and is used for further analysis.

Sample drug details from SIDER database

```
CID100000085    carnitine
CID100000119    gamma-aminobutyric
CID100000137    5-aminolevulinic
CID100000143    leucovorin
CID100000146    5-methyltetrahydrofolate
CID100000158    PGE2
CID100000159    prostacyclin
CID100000160    prostaglandin
CID100000175    acetate
CID100000187    acetylcholine
CID100000191    adenosine
CID100000206    glucose
CID100000214    PGE1
CID100000222    ammonia
CID100000232    arginine
CID100000244    benzyl
CID100000247    betaine
CID100000271    calcium
CID100000297    graphene
CID100000298    chloramphenicol
CID100000303    bile
CID100000305    choline
CID100000311    citric
CID100000312    chloride
CID100000338    salicylate
CID100000401    D-cycloserine
CID100000444    bupropion
CID100000450    estradiol
CID100000453    mannitol
CID100000564    EACA
```

Fig 6

CID100000085	A16AA01	CID100000119	L03AA03
CID100000119	N03AG03	CID100000137	L01XD04
CID100000143	V03AF03	CID100000143	V03AF04
CID100000143	V03AF06	CID100000158	G02AD02
CID100000159	B01AC09	CID100000160	G02AD01
CID100000175	G01AD02	CID100000175	S02AA10
CID100000187	S01EB09	CID100000191	C01EB10
CID100000191	J05AB03	CID100000191	S01AD06
CID100000206	A12AA07	CID100000206	A12BA01
CID100000206	A12CA01	CID100000206	B05CB01
CID100000206	B05CX01	CID100000206	B05XA01
CID100000206	B05XA03	CID100000206	B05XA07
CID100000206	G04BA03	CID100000206	V04CA02
CID100000206	V04CE01	CID100000206	V06DC01
CID100000214	C01EA01	CID100000214	G04BE01
CID100000232	B05XB01	CID100000247	A09AB02
CID100000247	A16AA06	CID100000271	A07XA03
CID100000271	A12AA20	CID100000298	D06AX02
CID100000298	D10AF03	CID100000298	G01AA05
CID100000298	J01BA01	CID100000298	S01AA01
CID100000298	S02AA01	CID100000298	S03AA08

Fig 7

Drug side effect details collected from SIDER database

CID100000085	CID000010917	C0000729	LLT	C0000729	Abdominal cramps
CID100000085	CID000010917	C0000729	PT	C0000737	Abdominal pain
CID100000085	CID000010917	C0000737	LLT	C0000737	Abdominal pain
CID100000085	CID000010917	C0000737	PT	C0687713	Gastrointestinal pain
CID100000085	CID000010917	C0000737	PT	C0000737	Abdominal pain
CID100000085	CID000010917	C0002418	LLT	C0002418	Amblyopia
CID100000085	CID000010917	C0002418	PT	C0002418	Amblyopia
CID100000085	CID000010917	C0002871	LLT	C0002871	Anaemia
CID100000085	CID000010917	C0002871	PT	C0002871	Anaemia
CID100000085	CID000010917	C0003123	LLT	C0003123	Anorexia
CID100000085	CID000010917	C0003123	PT	C0232462	Decreased appetite
CID100000085	CID000010917	C0003467	LLT	C0003467	Anxiety
CID100000085	CID000010917	C0003467	PT	C0003467	Anxiety
CID100000085	CID000010917	C0003811	LLT	C0003811	Arrhythmia
CID100000085	CID000010917	C0003811	PT	C0003811	Arrhythmia
CID100000085	CID000010917	C0004093	LLT	C0004093	Asthenia
CID100000085	CID000010917	C0004093	PT	C0004093	Asthenia
CID100000085	CID000010917	C0004238	LLT	C0004238	Atrial fibrillation
CID100000085	CID000010917	C0004238	PT	C0004238	Atrial fibrillation
CID100000085	CID000010917	C0004604	LLT	C0004604	Back pain
CID100000085	CID000010917	C0004604	PT	C0004604	Back pain
CID100000085	CID000010917	C0006277	LLT	C0006277	Bronchitis
CID100000085	CID000010917	C0006277	PT	C0006277	Bronchitis
CID100000085	CID000010917	C0007222	LLT	C0007222	Cardiovascular disorder
CID100000085	CID000010917	C0007222	PT	C0007222	Cardiovascular disorder
CID100000085	CID000010917	C0008031	LLT	C0008031	Chest pain
CID100000085	CID000010917	C0008031	PT	C0008031	Chest pain
CID100000085	CID000010917	C0009450	LLT	C0009450	Infection
CID100000085	CID000010917	C0009450	PT	C0009450	Infection
CID100000085	CID000010917	C0009806	LLT	C0009806	Constipation
CID100000085	CID000010917	C0009806	PT	C0009806	Constipation
CID100000085	CID000010917	C0010200	LLT	C0010200	Cough
CID100000085	CID000010917	C0010200	PT	C0010200	Cough
CID100000085	CID000010917	C0011570	LLT	C0011570	Depression
CID100000085	CID000010917	C0011570	PT	C0011570	Depression
CID100000085	CID000010917	C0011991	LLT	C0011991	Diarrhoea

Fig 8

Machine learning using SVM and Random Forest

The model was trained using SVM. The input data was split as 80% for training and 20% for testing. The classification criteria used is the adverse drug reactions. The classification is based upon the severity of the side-effect. For example: headache and abdominal pain are classified as “Low”, Depression, Back pain, Abdominal cramps are classified as “Medium” and Cardiovascular disorder, Anemia, Bronchytis are classified as “High”. Machine learning classification

techniques were used to predict the adverse side effects when polypharmacy was involved.

Accuracy measurement before using SMOTE

Technique	Accuracy %
SVM	50%
Random Forest	73%

Table 3

The analysis showed that Random Forest technique was able to predict the classification with greater accuracy compared to SVM. The SMOTE technique which was recommended in previous researches was applied here to balance the input dataset.

Accuracy measurement after using SMOTE

Technique	Accuracy %
SVM	67%
Random Forest	75%

Table 4

Drug-Drug interaction adverse effects measurement

Interacting Drug 1	Interacting Drug 2	side effects predicted	Correct (y/n)
Adenosine	Baclofen	Agitation	Yes
Baclofen	Atenolol	High Blood pressure	Yes
Atenolol	Caffeine	Bronchospasm	Yes
Caffeine	Amiodarone	None	No

Table 5

Conclusion

Adverse drug-drug interaction is a costly affair on human health. It is commonly known after the drugs are in the market and when patients report complaints and doctors detect health hazards. Most of the time it is during post-prescription stage after the patient has consumed the medicine. The damage is already done to the patient. In this study we have tried to come up with a preventive measure and a means to detect if the combination of the medicines when taken by a patient will

impact his/her health adversely or not. We have seen that the prediction model and the current stage of machine learning is at an accuracy level of 75%. The SMOTE technique was helpful in making the input dataset as a balance one. Random forest classification method was able to provide greater accuracy than SVM. We were also able to compare the results of the prediction with the data in the Sider database and see that there is 75% near accuracy. Definitely, there is scope for improving this very much as was evident during the study. We further intend to improvise the accuracy of this prediction by combining the best of all classification algorithms and also extend it to include the health record of the patient as an additional parameter as well.

References

1. A. Pradhan, ``Support vector machine_A survey,'' *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 8, pp. 82_85, Aug. 2012.
2. Adoption of Electronic Health Records: A roadmap for India https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5116537/
3. Alamsyah, T., Marianthi, D., Hayati, W., & Usrina, N. (2021). Drug user behavior about the development and rehabilitation process in Banda Aceh correctional institution. *International Journal of Health & Medical Sciences*, *4*(1), 88-94. https://doi.org/10.31295/ijhms.v4n1.1348
4. Big data in healthcare – the promises, challenges and opportunities from a research perspective: A case study with a model database https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977694/(2016)
5. BIG Data revolution in healthcare sector: http://www.aircconline.com/ijist/V6N2/6216ijist16.pdf
6. Challenges with EMRs: https://www.healthcareitnews.com/news/biggest-ehr-challenges-2018-security-interoperability-clinician-burnout
7. Clinical trials https://www.clinicaltrials.gov/ct2/show/NCT04011124
8. D.-S. Cao, Q.-N. Hu, Q.-S. Xu, Y.-N. Yang, J.-C. Zhao, H.-M. Lu, L.-X. Zhang, and Y.-Z. Liang, ``In silico classi_cation of human maximum recommended daily dose based on modi_ed random forest and substructure _ngerprint,'' *Anal. Chim. Acta*, vol. 692, nos. 1_2, pp. 50_56, Apr. 2011
9. Drug designing open access: https://www.longdom.org/scholarly/structure-based-drug-design-journals-articles-ppts-list-707.html
10. Drug-target interaction prediction based on drug fingerprint information and protein sequence, Yang Li, Yu-An Huang, Zhu-Hong You, Li-Ping Li and Zheng Wang, https://www.researchgate.net/publication/335246878_Drug-Target_Interaction_Prediction_Based_on_Drug_Fingerprint_Information_and_Protein_Sequence, Aug 2019
11. E. Lounkine, M. J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon,J. L. Jenkins, P. Lavan, E. Weber, A. K. Doak, S. Côté, B. K.Shoichet, and L. Urban, ``Large-scale prediction and testing of drug activity on side-effect targets,'' *Nature*, vol. 486, no. 7403, pp. 361_367,Jun. 2012.
12. EHR adoption in India: www.indjst.org/index.php/indjst/article/download/100211/73180
13. Electronic Medical Records: https://www.usfhealthonline.com/resources/key-concepts/what-are-electronic-medical-records-emr/
14. G. Chandrashekar and F. Sahin, ``A survey on feature selection

methods,'*Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16_28, Jan. 2014.
15. https://en.wikipedia.org/wiki/Adverse_drug_reaction
16. https://en.wikipedia.org/wiki/Ligand
17. IBM research Health Informatics: https://www.research.ibm.com/haifa/dept/vst/analytics.shtml
18. Identifying children and saving lives one thumbprint at a time (2016) : https://msutoday.msu.edu/news/2016/identifying-children-and-saving-lives-one-thumbprint-at-a-time/
19. Improving the Use of Big Data Analytics within Electronic Health Records: https://www.sciencedirect.com/science/article/pii/S1877050918301091
20. K. M. Giacomini, R. M. Krauss, D. M. Roden, M. Eichelbaum, M. R. Hayden, and Y. Nakamura, ``When good drugs go bad,'' *Nature*,vol. 446, no. 7139, pp. 975_977, Apr. 2007.
21. L. Breiman, ``Random forests,'' *Mach. Learn.*, vol. 45, no. 1, pp. 5_32,Oct. 2001.
22. M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, ``The SIDER database of drugs and side effects,'' *Nucleic Acids Res.*, vol. 44, no. D1, pp. D1075_D1079, Oct. 2015
23. Ndoloe, L. A., & Rumlaklak, N. D. (2018). Application design in determining the shortest route using tabu search algorithm. *International Journal of Physical Sciences and Engineering*, *2*(3), 77–91. https://doi.org/10.29332/ijpse.v2n3.233
24. O. Ursu, J. Holmes, C. G. Bologa, J. J. Yang, S. L. Mathias, V. Stathias, D.-T. Nguyen, S. Schürer, and T. Oprea, ``DrugCentral 2018: An update,'' *Nucleic Acids Res.*, vol. 47, no. D1, pp. D963_D970, Oct. 2018.
25. Pharma sciences https://www.nuventra.com/services/pharmacokinetics-pharmacodynamics/drug-drug-interaction-studies/
26. Predicting drug interaction with Adenosine Receptors using machine learning and SMOTE: Abdel Rahman I. SAAD, Yasser MK Omar, Fahima A Maghraby, IEEE DOI 10.1109/Access.2019.2946314,
27. S. F. Abdoh, M. A. Rizka, and F. A. Maghraby, ``Cervical cancer diagnosis using random forest classi_er with SMOTE and feature reduction techniques,'' *IEEE Access*, vol. 6, pp. 59475_59485, 2018
28. S. Gessi, S. Merighi, P. A. Borea, S. Cohen, and P. Fishman, ``Adenosine receptors and current opportunities to treat cancer,'' in *The Adenosine Receptors*. Cham, Switzerland: Humana Press, 2018, pp. 543_555. https://link.springer.com/chapter/10.1007/978-3-319-90808-3_23
29. Solutions for healthcare and sciences: https://databricks.com/solutions/industries/healthcare-life-sciences
30. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). Get vaccinated when it is your turn and follow the local guidelines. *International Journal of Health Sciences*, *5*(3), x-xv. https://doi.org/10.53730/ijhs.v5n3.2938