A Comprehensive Review and Application of Interpretable Deep Learning Model for ADR Prediction

Shiksha Alok Dubey[1], Anala A. Pandit[2]
Department of Computer Application
Veermata Jijabai Technological Institute (V.J.T.I)
Mumbai, India

***Abstract*—Drug safety is a pressing need in today's healthcare. Minimizing drug toxicity and improving the individual's health and society is the key objective of the healthcare domain. Drugs are clinically tested in laboratories before marketing as medicines. However, the unintended and harmful effects of drugs are called Adverse Drug Reactions (ADRs). The impact of ADRs can range from mild discomfort to more severe health hazards leading to hospitalization and in some cases death. Therefore, the objective of this research paper is to design a framework based on which research papers are collected from both ADR detection and prediction domain. Around 172 research articles are collected from the sites like ResearchGate, PubMed, etc. After applying the elimination criteria the author categorized them into ADR detection and prediction themes. Further, common data sources and algorithms as well as the evaluation metrics were analyzed and their contribution to their respective domains is stated in terms of percentages. A deep learning framework is also designed and implemented based on the research gaps identified in the existing ADR detection and prediction models. The performance of the deep learning model with two hidden layers was found to be optimum for ADR prediction and further, the non-interpretability part of the model is addressed using a global surrogate model. The proposed architecture has successfully addressed multiple limitations of existing models and also highlights the importance of early detection & prediction of adverse drug reactions in the healthcare industry.**

Keywords*—Drug safety; adverse drug reactions; early detection; deep learning; interpretable models

I. INTRODUCTION

Drug safety is a pressing need of today's healthcare. Minimizing drug toxicity and improving the health of individuals and society is the key objective of the healthcare domain. The drug development process starting from discovery to market is long and costly. Rigorous efforts are involved in clinical trials to ensure the safety and efficacy of the developed drugs. Clinical trials are performed on any new drug substance to check its safety and effectiveness against the particular disease before marketing them as medicines to the general population [1]. Currently, all developed drugs have risks associated with them [2] and only those drugs whose curative impact is greater than the risk, are marketed as medicines. Any unwanted, undesired effects of drugs on human health are considered Adverse Drug Reactions (ADRs). According to the definition provided by WHO (World Health Organization), an ADR can be unintended and occurs at doses normally used in man for the prophylaxis, diagnosis or therapy of disease, or modification of physiological function" [3]. Simply, it can be seen as an unpleasant or unexpected effect of drugs on patients. The impact of ADRs is considered to be one of the reasons behind mortality and morbidity in humans. The contribution of ADRs is about 5% of all hospital admission and it is considered the fifth most common reason for mortality during hospitalizations [4]. The severity and harmfulness of the reported ADRs have caused a ban on many developed drugs. In about 20 years around 40 drugs are withdrawn from the drug market due to the severe reactions caused due to them [5]. Around 50% are banned from the US market [6] then Germany [5] and finally from the Europe drugs market. The most common occurring toxicities due to drugs are cardiotoxicity (32%, [13]), hepatotoxicity (20%, [8]) then death risk (10%, [4]), and finally risk of overdose (7%, [3]). A recent example of a Sibutramine (Meridia) drug got initial permission from FDA to be sold as an appetite suppressant, but in 2010 it was banned from the market as it caused an increase in heart disease and heart stroke risk in patients. The severity of ADRs can also be measured in terms of the burden of healthcare cost and length of hospital stay [7]. A variety of factors are also responsible for the development of ADRs in humans. These factors can be classified as patient-related, drug-related, and social environment-related [8]. Gender and age are critical patient-related parameters that need to be considered while assessing the impact of ADRs on individuals while drug dosage and drug-drug interaction are important drug-related factors [8]. Confounding factors like smoking and alcoholism are crucial in the development of ADRs [8]. For better patient safety and improving healthcare, it is important to not only predict an ADR on time but also detect it at an early stage.

The following diagram Fig. 1, shown illustrates that ADRs are included as part of ADE (Adverse Drug Event) which is again a subset of adverse events. But the fact that separates ADRs from ADE is that they are caused due to drug intake even at normal dosage.

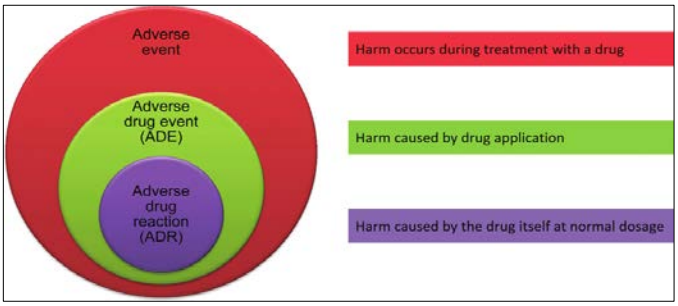


Fig. 1. Classification of Adverse Events. Adverse Events Include all Harmful Events Occurring during Treatment with a Drug without the Necessity of a Causal Link between the Drug and the Reaction. If the use of Medication is Causal to the Reaction, the Condition is called an Adverse Drug Event. A Subform of Adverse Drug Events is Adverse Drug Reactions that are Triggered by the Drug Itself Despite its Appropriate Dosage [9].

II. Framework for Research Paper Selection

In the last two decades, a lot of research has been done in this field of ADR identification and improving drug safety. Researchers have published their research works highlighting the need for the detection and prediction of ADRs in the healthcare industry. Therefore the purpose is to first summarize these published research articles from multiple perspectives and then apply deep learning models for ADR prediction. The research papers are collected from both domains. Depending on the elimination criteria defined, only the relevant research works are selected for further analysis. Arksey and O'Malley's [10, 11] methodological framework is used for selecting research papers for literature review. This framework has helped researchers to concentrate on a single domain for a short duration and identify research gaps depending on the collected research works. The entire methodology can be summarized as follows:

Stage 1: Identifying the research question

As previously discussed this review focuses mainly on the research studies done in the past related to the detection and accurate prediction of ADRs. Detecting an ADR from data is important before predicting it, therefore research papers are included from both domains.

Theme 1: ADR Detection

What makes ADR detection critical for drug safety? What is the different ADR signal detection techniques applied to datasets? How the different techniques are evaluated on a variety of datasets?

Detecting an ADR is an important step to improve healthcare and drug safety [12]. It is important to detect ADR and distinguish it from the symptoms of the disease. Different detection techniques are defined for different datasets.

Theme 2: ADR Prediction

Why accurate prediction of ADR is important for better patient safety and minimizing ADR occurrences? What is the different prediction models applied for ADR prediction? How computational models are useful in preventing severe ADRs in the future?

Predicting an ADR can prevent its occurrence and minimize healthcare costs [13]. Different models have been applied in the past, present, and future to predict and prevent such ADRs. The extent of this review study includes machine learning and deep learning models for ADR prediction [14].

Stage 2: Collecting the research studies

As previously discussed, the author has collected research articles related to ADR detection and prediction domain published throughout 10yrs. The research studies are from both computer science and biomedical domain. Major search engines and databases from where these publications & databases used in those publications are:-

PubMed:-It is a search engine that provides easy access to the MEDLINE database [15] and is freely available. It also provides access to abstracts and references related to biomedical as well as life science domains [16].

ResearchGate:- It is a European social networking site [17] that provides a common platform for both researchers and scientists. The majority of research articles related to different domains are published on ResearchGate [18] for access to both researchers and academic professionals.

The indexing mechanism available in PubMed is Medical Subject Headings (MeSH) [19].

MeSH:- It is a controlled comprehensive vocabulary [19] used for indexing journals available on PubMed. This indexing is very helpful for searching research articles and journal papers. The research studies were searched using different keywords related to 'adverse drug reaction', 'prediction related ADR', and 'detection of ADR', and different datasets were openly accessible and acquired through ethical permission.

Query-based search: - The different query strings related to pharmacovigilance [20] are used for searching different articles on Google Scholar. The articles are searched based on heading, abstract and main content.

Stage 3: Select only the relevant studies.

The author has defined some criteria based on which only the relevant research studies were selected. The elimination criteria are listed below:-

Duplicate research papers are eliminated.

Research studies not related to the review.

The research papers largely focused on the biomedical domain.

The research studies were more related to clinical research.

The research article more focused on drug-drug interaction and the genetic interaction of drugs.

Other unrelated research works.

After filtering, only relevant research works are selected for further analysis, and the results drawn are presented in this paper.

Stage 4: Charting the Data

The author has reviewed the research papers from multiple aspects. The various perspectives based on which the research studies are evaluated are described below:-

Search Engine/Database

Year of Publication

Journal/Conference

Name of the research paper

Datasets used

Models applied for ADR detection & prediction

Drugs mentioned for a given ADR

Evaluation metrics applied

Stage 5: Summarizing and reporting results

The research studies are summarized and segregated based on the approach used. In the initial phase, only the relevant studies are considered and the irrelevant ones are filtered out. Then the research works are grouped according to themes 1 and 2. Theme 1 is ADR detection whereas theme 2 is ADR prediction. The layout of the entire process is outlined in the flowchart shown in Fig. 2.

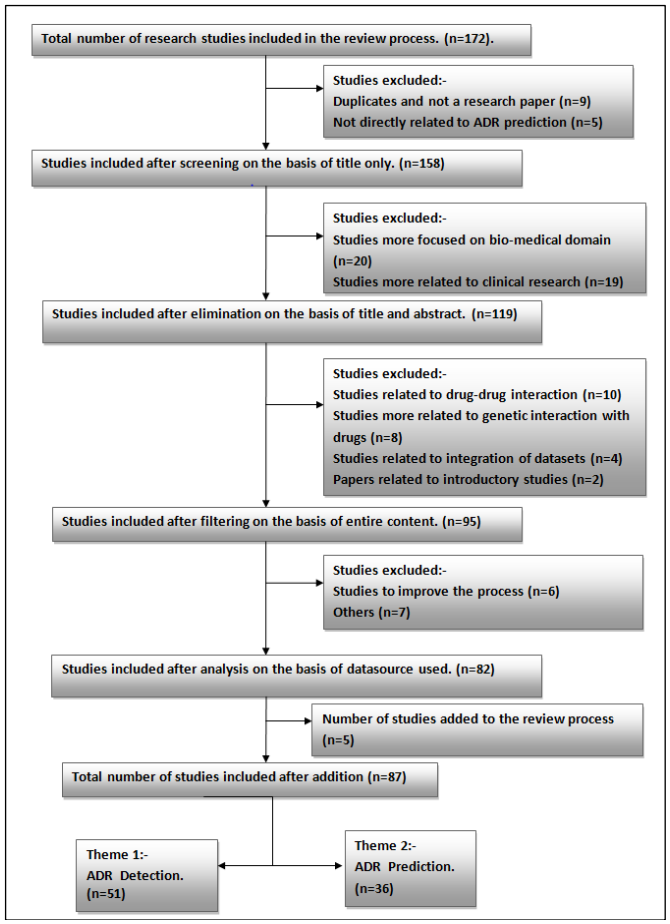


Fig. 2. Framework for Research Paper Selection.

The layout of the entire process is illustrated in the above diagram. Around 172 research works are collected showing the adverse effects and reactions of drugs on human health. The elimination criteria are used to eliminate irrelevant research papers. Repeated research works (n=9) and studies not related to ADR prediction (n=5) are filtered. Around 158 research works remained after elimination. About 19 research studies were eliminated in the screening process that belongs to the clinical research domain. Around 20 research works, more related to the biomedical domain were also eliminated. After filtering it only 119 research works remained. Research related to drug-drug interaction (n=10) and genetic interaction of drugs (n=8) were also screened out. Introductory studies (n=2) and research works related to the integration of datasets (n=4) are eliminated. A total of 95 research papers remained. Some research studies related to the improvement of ADR detection and prediction process (n=6) along with others (n=7) are also eliminated. In addition, five more research papers were made for the final review analysis. Finally, after filtering the research studies based on elimination criteria previously defined, the author identified about 87 research papers for further analysis.

Classifying them according to the two themes of ADR detection and prediction, there are about 51 papers associated with ADR detection and the remaining 36 are related to ADR prediction.

III. Summary of ADR Datasets

Incidents of adverse reactions have been in existence for more than two decades. Over the period many countries have established pharmacovigilance centers [21] for collecting the reported occurrences of ADRs from medical practitioners and healthcare workers. These centers contribute to the postmarketing surveillance of ADRs. Many secondary data sources have been established by collecting both prescription data and ADR information. ADRs are also monitored actively through clinical trials and identified in different structured and unstructured data sources [22]. Different ADR-related data sources are listed and discussed in Table I.

TABLE I. Summary of ADR Datasets

ADR related datasources	Description	Website
Primary databases ***Spontaneous reporting Systems(SRS)***		
FAERS[23] EMA[24] UMC[25]	The incidents of ADRs are reported to the regulatory bodies of the country. They are analyzed and stored in databases for further action against the reported drug. These databases are also available for review and research process.	https://open.fda.gov/data/faers/ https://www.ema.europa.eu/en https://www.who-umc.org/
Electronic Health Records[26]	This database contains records of patients admitted to the hospital. The datasource is very accurate as it records all information about the patient's condition the disease and its recovery phases.	This database can only be required through ethical permission from the required regulatory authority
Clinical narratives	The narratives and discharge summaries are written by experienced healthcare professionals. It contains	This data again requires ethical permission for

	data about a patient, disease, prescription information, and the treatment given. This information is very accurate and precise.	using it as part of the research.
Major secondary ADR databases		
SIDER 4.1[27] (Medical Literature)	This dataset includes data of 1430 marketed medicines and their recorded 5868 ADRs. It also includes around 139756 drug-SE association pairs.	http://sideeffects.embl.de/
OFF-SIDES[28]	Offsides is a database of drug side effects that were found, but are not listed on the official FDA label.	http://tatonettilab.org/resources/nsides/
TWOSIDES [28]	An online available dataset containing information about drug-drug interaction and side-effects due to drug-drug interactions.	http://tatonettilab.org/resources/nsides/
ADReCS [29]	A comprehensive ADR ontology database.	http://bioinf.xmu.edu.cn/ADReCS
Medical Forums	These are public websites used for posting health-related inquiries.	https://www.dailystrength.org/
Major API(Active Pharmaceutical Ingredients) interaction databases		
DrugBank [30]	Comprehensive online database containing information on drugs & drug targets.	https://go.drugbank.com/
PubChem [31]	A resource with information on chemical substances and their biological activities	https://pubchem.ncbi.nlm.nih.gov/
SuperDRUG 2[32]	SuperDRUG2 is a unique, one-stop resource for marketed and approved drugs containing 4,600 active pharmaceutical ingredients [32].	http://cheminfo.charite.de/superdrug2/
SuperTarget	A resource that contains information about drug and target proteins and analyses their associations.	https://bioinformatics.charite.de/supertarget/
STITCH[33]	A resource collecting known and predicted interactions between chemicals and proteins.	http://stitch.embl.de/
PharmGKB [34]	The pharamakogenomic knowledgebase is a publicly available online knowledge base used for aggregation and integration of information on drugs and analyzing their impact on genetic variation.	https://www.pharmgkb.org/
KEGG(Kyota Encyclopedia of Genes & Genomes)[35] & GO(Gene Ontology) [36]	It is a collection of databases dealing with genomes, biological pathways, diseases, drugs, and chemical substances. GO resource contains information about gene function.	https://www.genome.jp/kegg/ http://geneontology.org/

Different related data sources are grouped into separate categories. The basic categories defined are primary and secondary data sources. The table incorporates a variety of ADR-related data resources for both ADR detection and prediction.

A. ADR Detection

The research studies are majorly done in USA (n=33/51, 65%), Europe (n=6/51, 12%) and Korea (33/51, 8%). Apart from this research contributions from Australia (n=3/51,6%) and India (n=2/51,4%) are also considered.

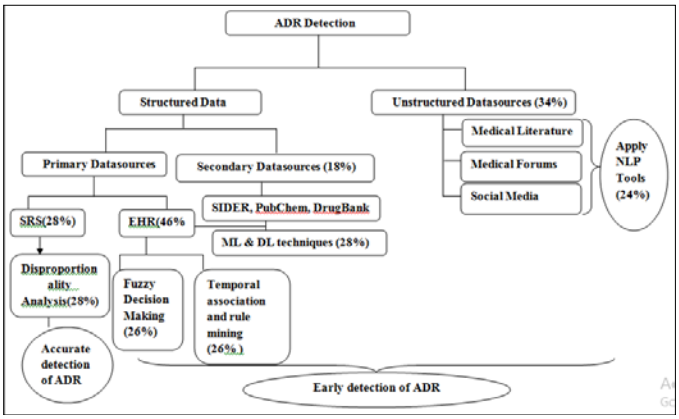


Fig. 3. The Layout of the ADR Detection Process.

As shown in Fig. 3, the ADR detection data sources are categorized into two groups that are structured and unstructured datasets. The unstructured dataset is used in about 34% (n=17/51) of the research works while structured datasets are again categorized into primary and secondary data sources. The primary data sources are again divided into SRS and EHR. SRS is utilized in around 28% (n=14/51) of the research papers while EHR is in 46% (n=23/51) of the research. The secondary data source includes information about the drug-ADR association and is included in about 18% (n=9/51) of the research works. The different techniques are applied based on the data sources used.

DPA (Disproportionality analysis) is applied in around 28% of research papers where SRS is involved to validate the potential drug-ADR association.

Fuzzy Decision Making & Temporal Association Mining is applied equally in about 26% of the research studies for the early detection of ADRs.

Machine Learning (ML) & Deep Learning (DL) models are applied for secondary and primary data sources in around 28% of the research studies. The models are trained to detect unknown drug-ADR associations from datasets.

Finally, NLP (Natural Language Processing Tools & Techniques) are applied in about 24% of the research studies for extracting meaningful insights from unstructured text.

EHR and unstructured text has been used to early detect an ADR while SRS is helpful in the accurate detection of ADR.

B. ADR Prediction

The geographical research distribution for ADR prediction shows that the majority of research work is carried out in the USA at 44 %(n=16/36) followed by China at 22 %(n=8/36) and finally in Europe at 17 %(n=6/36). Other countries like Croatia, Romania, India, Israel, Iran, Korea, and Japan have also contributed to the research in the ADR prediction domain. The major steps performed for ADR prediction are illustrated as follows:-

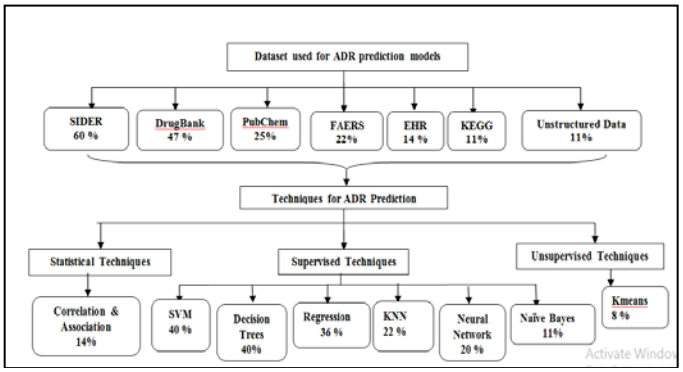


Fig. 4. The Layout of the ADR Prediction Process.

The Fig. 4, illustrates that the datasets mainly used for ADR prediction are SIDER 60%(n=21/36), Drugbank 47% (n=17/36), PubChem 25% (n=9/36) and FAERS 22%(n=8/36). Further, the techniques applied for ADR prediction are divided into three categories that are statistical, supervised, and unsupervised techniques. The statistical methods are further defined as correlation and association methods that contribute to 14% (n=5/36) of the research works while unsupervised techniques are further classified as Kmeans are applied in about 8%(n=3/36) of the research works. The common supervised techniques applied in the research works are SVM (Support Vector Machine) 40%(n=14/36), Decision Trees 40% (n=14/36), Regression techniques 36 % (n=13/36), KNN (K-Nearest Neighbor) 22% (n=8/36) and Neural Network 20 %(n=7/36).

The models are also analyzed based on evaluation metrics applied to the models for examining their performance.

The diagram in Fig. 5, depicts the percentage contribution of different evaluation metrics to ADR detection & prediction models. The precision & recall evaluation metric contributes to about 48% of ADR detection research papers while 60% of ADR prediction research works. The specificity, sensitivity & AUC are applied in about 30% of ADR detection research studies while only AUC is applied in about 70% of ADR prediction research papers. Lastly, the accuracy metric is used in around 40% of the research studies while the Ranking of drug-ADR association based on different metrics is involved in about 20% of the research works.

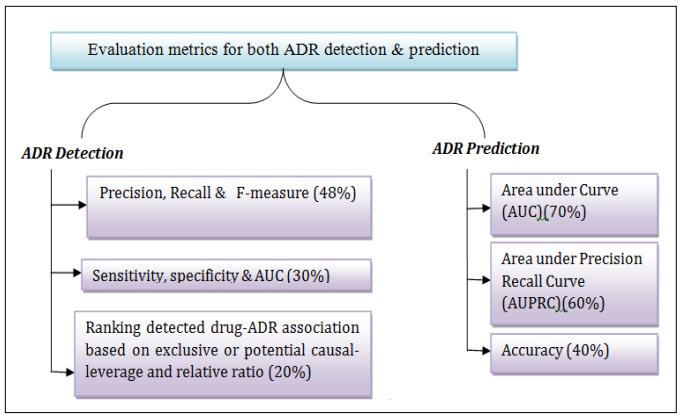


Fig. 5. Evaluation Metrics for ADR Detection and Prediction.

C. Research Gaps Analysis

After reviewing the research studies, the author has identified some major dataset and technique-related limitations that are shown in the diagram:-

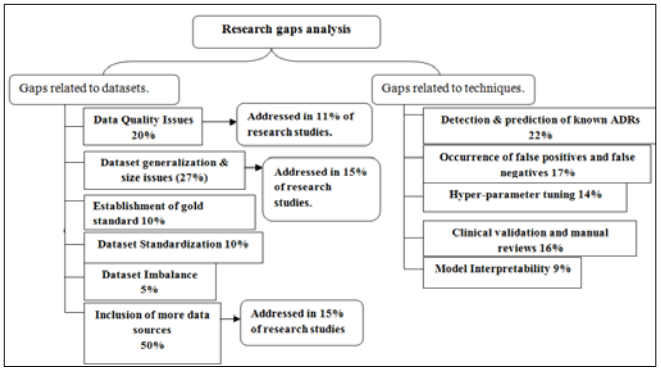


Fig. 6. Research Gap Analysis.

The Fig. 6, shows a thorough gap analysis depending on selected research papers. The significant research gaps related to ADR datasets are data quality, data generalization, and integration of more data sources which is specified in about 50% of the research papers. The research gaps are also analyzed based on the techniques applied for detecting and predicting an ADR. The major gaps discussed are the detection & prediction of known ADRs, the occurrence of false positives and false negatives, hyper-parameter tuning, and clinical validation. We have tried to address some research gaps in our research work but still many needs to be addressed for the future research study. These limitations form the basis to design our model for ADR prediction.

A framework is developed based on the gap analysis illustrated in Fig. 7.

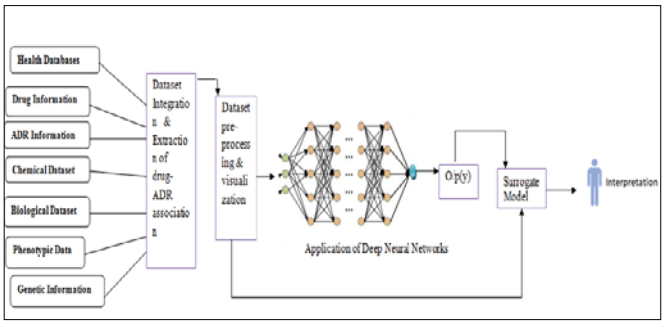


Fig. 7. Framework of the Proposed Model.

The framework signifies the key steps performed for ADR prediction. It tries to address the above-stated research gaps namely inclusion of more data sources, dataset imbalance, dataset size issues, and detection & prediction of known ADRs. The steps shown in the framework are practically implemented and results are derived accordingly.

IV. Dataset Selection and Approaches for Integration

The FAERS [23] data source used as input is a primary data source. It is available and freely accessible online. The data is collected and stored through an authentic process and

validated. This dataset is presented both in ASCII and CSV format. Around three million records were collected from the FAERS dataset dated from 2019 to 2020 end in ASCII format. Once downloaded and extracted the overall dataset is visualized in Fig. 8.

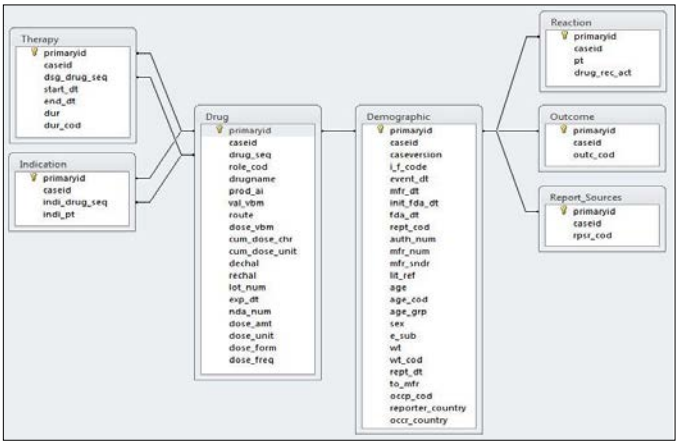


Fig. 8. FAERS Dataset [23].

The data in the FAERS dataset is unreadable and segregated across multiple tables. Therefore it is required to convert the dataset into a readable format and then integrate it using primaryid and caseid. The size of the integrated dataset is huge therefore it is necessary to detect and identify significant drug-ADR associations. The disproportionality analysis technique is applied for extracting such associations from the dataset.

The drug-ADR association is calculated in terms of PRR (Proportionality Reporting Ratio) [37]. Only those associations which are greater than the threshold value i.e. PRR>=3 are filtered for further processing.

The output of the ADR detection algorithm is illustrated in Table II:-

TABLE II. RESULTS OF ADR DETECTION

Product	Adverse Event	Count	p_value	PRR
cc-10004	lymph node tuberculosis	102	-6167.969382	1.04847E+13
rifampicin	skin papilloma	6	-6168.662469	5.24282E+12
rapamune	hemiplegia	8	-6168.662467	5.24279E+12
alpelisib	osteonecrosis of jaw	16	-6168.662472	5.24272E+12

Overall the result of the initial experiments provides us with a processed and filtered FAERS dataset which is used further for integration with other data sources. The final ADR prediction is performed using drug characteristics as well as patient characteristics.

SIDER contains information regarding the marketed medicines along with their recorded ADRs. This dataset is secondary and is easily available on the internet for research purposes. The data source also includes information about drug indications on patients which are extracted from unstructured text using NLP tools and techniques [27]. These drug indications help to distinguish ADRs from symptoms of disease and thus reducing the number of false positives. It is one of the most popular datasets used in ADR detection & prediction-based research study. It has been used in almost 60% of the research work done.

DrugBank was created by the University of Alberta and The Metabolomics Innovation Centre in Alberta, Canada [30]. It is a comprehensive, easily available, online data source that includes data about drugs and the protein targets of drugs. It also includes the components of proteins in terms of enzymes, transporters, receptors, and ion channels. The biological effect of drugs in terms of drug toxicity is also included as part of this research dataset. This dataset was acquired after obtaining the required permission from the authorities and assuring them of its ethical use.

PubChem includes information about drug molecules along with their chemical composition and their effect in response to the biology of patients. This data source is developed by NCBI (National Center for Biotechnology Information) which is a part of NLM (National Library of Medicine). The NLM is also included as a part of NIH (National Institutes of Health) of the USA [31]. It is also easily available for research purposes online.

The selected datasets are integrated using two different techniques. Each technique is illustrated in the following Fig. 9 and 10.

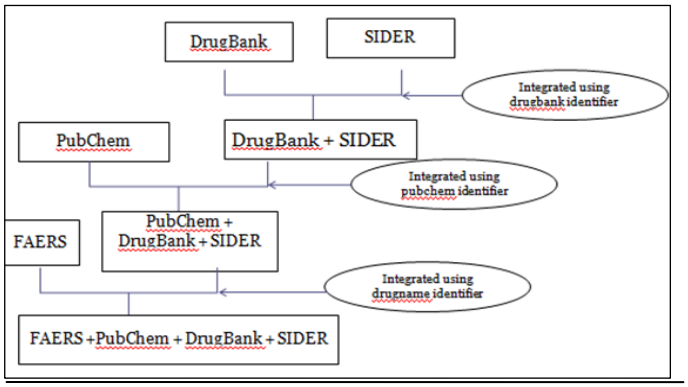


Fig. 9. Drug Identifier-based Integration.

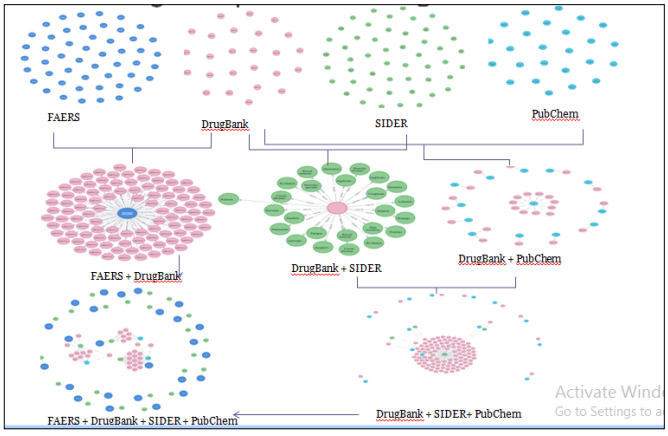


Fig. 10. Knowledge Graph Integration [38].

In drug identifier-based integration the DrugBank and SIDER are integrated using the drugbank identifier. Further PubChem is integrated using the PubChem identifier and lastly, FAERS is included using drug names. Similarly, in the case of knowledge graph [38] based integration, knowledge graphs are constructed using nodes of different datasets like drugs, target proteins, enzymes, pathways, indications, and adverse drug reactions. In the above figure, knowledge-graph integration information is derived from knowledge graphs which are used for identifying side-effects as well as detecting probable ADR for the prescribed medicines.

Further, the features of the integrated datasets are reviewed by the domain expert, and useful feedback and inputs were obtained by the author accordingly. Some features were dropped from the dataset while some were added as per their recommendations. The feature variables included as part of the integrated dataset are the type of target and target sequence. The type of targets can be divided into four categories receptors, ion channels, enzymes, and carrier molecules. Target sequences are genetic variants targeted by a given drug molecule. Apart from target type and target sequence some other features were also added as part of this dataset which is described in Table III.

TABLE III. FEATURE VARIABLE DESCRIPTION

Feature Variable	Description
LogP	Lipophilicity is a valuable parameter of the drug which affects its activity in the human body. The Log P value of the compound indicates the permeability of the drugs to reach the target tissue in the body[39]
LogS	The aqueous solubility of a compound significantly affects its absorption and distribution characteristics. Typically, a low solubility goes along with a bad absorption, and therefore the general aim is to avoid poorly soluble compounds. Our estimated logS value is a unit stripped logarithm (base 10) of the solubility measured in mol/liter.[39]
CYP inhibitors	The inhibitors are responsible for delaying the action of target proteins and that lead to a large amount of drug disposition in the human body which is harmful and severe.
Toxicity	The toxic nature of the drug molecule on the human body.

A. Dataset Preprocessing

The integrated dataset contains several redundant columns, null values, and categorical feature variables that must be pre-processed before further analysis. The steps involved in the pre-processing of both identifiers integrated and knowledge graph integrated datasets are described in the following Fig. 11.

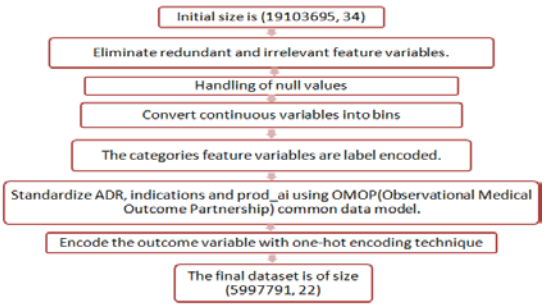
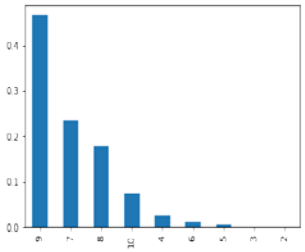


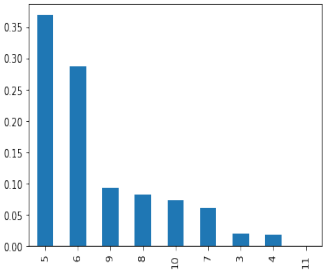
Fig. 11. Pre-processing Steps on Integrated Dataset.

The data distribution of the significant feature variables in the integrated dataset is visualized in the following bar charts shown in Table IV.

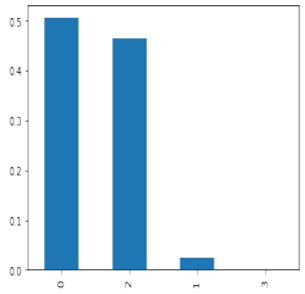
TABLE IV. DISTRIBUTION OF FEATURE VARIABLES



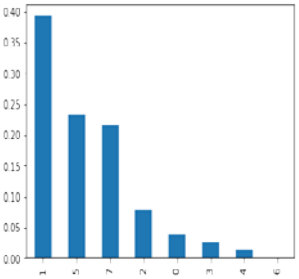
Age
The age feature variable demonstrates that the majority of people who experience ADRs belong to the category of Aged, Adults, and middle-aged. It is in the range 44 years to 64 yrs. and above 79yrs.



wt
The weights of the people experiencing ADR belong in the range of 50 to 60 kg and above 80kg.



Sex
The ADRs are mainly reported by the female population in the country followed by males.



Hepatotoxicity
The most occurring toxic effects on the liver are identified as
B (likely cause of clinically apparent liver injury)
E (unlikely to be a cause of clinically apparent liver injury)and
A (well-known cause of clinically app arent liver injury).
It ranges from severe to mild toxicity

B. Kruskal Wallis Test

This test is used to determine whether or not there is a statistically significant difference between the medians of three or more independent groups [40]. The result of this test is shown in the output below.

```
H-statistic:    84253398.14222576
P-value:        0.0
Reject NULL hypothesis - Significant differences exist between groups.
```

The p-value is zero which is less than 0.05 which shows that a significant difference exists between groups and rejects the NULL hypothesis.

C. ADR Prediction Dataset

The current dataset includes only positive ADR samples. For any prediction problem, a balance of positive and negative data samples is required. Therefore the author applies GANs (Generative Adversarial Networks) [41] architectures to the original dataset and generates negative data samples based on the features of the original dataset. The output of the application of GANs is shown in Fig. 12.

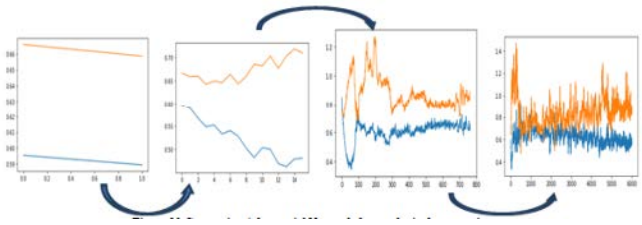


Fig. 12. Generative Adversarial Network for Data Creation.

A combined dataset of 20 lakh records was generated for both presence and absence of ADRs. This dataset will be used for the implementation of ADR prediction algorithms. The target class distribution before and after the application of GANs is illustrated in Fig. 13.

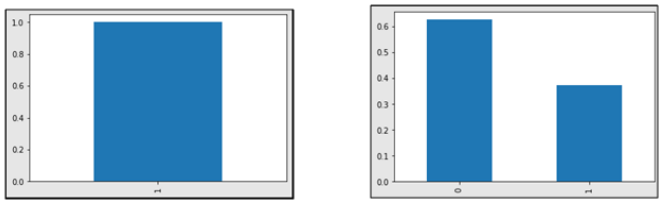


Fig. 13. Target Class Distribution.

The target class distribution shows the balance of positive and negative samples in the ADR dataset after the application of the GAN model.

V. Results

Deep learning models are Deep Neural Networks (DNN) containing non-linear processing units that transform raw data into higher-level representative information [40]. In recent years these techniques are actively applied in the field of drug discovery, precision medicine, protein engineering, genetic expression data analysis, and pharmacodynamics modeling [42]. Given the significant contribution of deep learning techniques in the domain of drug discovery, its capability can be very well extended to predict adverse reactions to drugs in humans. As previously discussed the ADR data sources both structured and unstructured are huge, diverse, and heterogeneous. DNN can successfully be applied to these data sources without the need for manual tuning. The initial training using a deep neural network is very complex and time-consuming but the network improves its performance by learning from input data.

Therefore the author proposes to apply deep learning models to the integrated dataset and evaluate its performance in terms of different evaluation metrics like accuracy, precision, recall, and F1 score. The model training is performed based on drug-ADR associations and other associated information. Deep learning performs well on a huge dataset. It also eliminates the need for hyper-parameter tuning. The number of hidden layers is optimized to provide the best results on the given dataset. The results obtained are shown in Table V:-

TABLE V. Deep Neural Network Results

	Accuracy	Precision	Recall	F1 Score
The model with one hidden layer	0.57	0.9	0.57	0.64
The model with two hidden layer	0.91	0.92	0.91	0.91
The model with three hidden layer	0.55	1	0.55	0.71

From the results obtained it can be observed that the performance of the model with two hidden layers provides is optimum for all evaluation metrics. The performance seems to be consistent for precision but it variates significantly for the other evaluation metrics. It gives poor performance for one and three hidden layers but the best performance for two hidden layers. The results can be visualized in below Fig. 14.

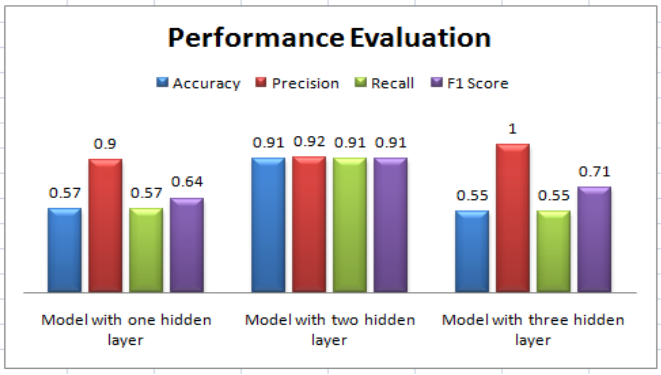


Fig. 14. Results Obtained by Models based on different Numbers of Hidden Layers.

Fig. 14 reflects the results obtained in the table and the performance of all evaluation metrics is consistent for the model with two hidden layers. It can be observed that the benefits of deep learning approaches are extensive but they suffer from the issue of non-interpretability. The 'black box' nature of deep learning techniques has restricted the interpretability of the model. The author has demonstrated the need for interpretable models for overall acceptability in the medical domain. Therefore to address this limitation the author has proposed the application of LIME (Local interpretable model-agnostic explanations) [43] for model explainability.

LIME is a technique that approximates any black-box learning model with a local, interpretable model to explain each prediction. From the definition it can be understood that LIME provides approximate explanations to individual prediction instances i.e. it is a local surrogate model. But to interpret the results based on the entire dataset SP-LIME [43] is applied. SP-LIME (Sub-modular Pick- Local Interpretable Model-Agnostic Explanation) tries to provide an answer to the question of developing trust for a given model for its acceptance. The trust is developed by dividing a given problem into several sub-problems for optimization. That means it identifies a series of instances along with their predictions that reflects the overall performance of the model based on the given data. The instances are selected in such a

manner that the features which are responsible for explaining different predictions are given higher importance value.

The results obtained by applying the SP-LIME algorithm to the given dataset are shown as follows:-

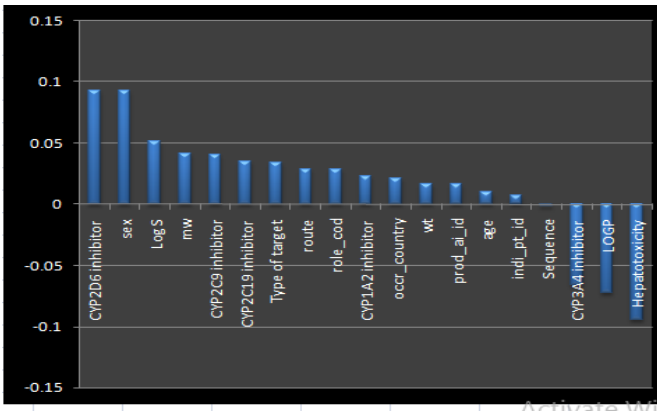


Fig. 15. SP-LIME Results.

Fig. 15 describes that the feature variable CYP2D6 inhibitor and sex contribute the highest to the target outcome prediction positively while hepatotoxicity, LOGP, and CYP3A4 are responsible for predicting the ADR outcome negatively. The proposed framework tries to address the research gaps stated in the existing research works. The inclusion of more data sources is identified in about 50% of the research studies, this issue is tackled by our proposed model. Other than this the model is trained on only validated drug-ADR association which is stated in about 16% of the research work. Lastly, the model's interpretability issue is also handled using a surrogate model. Therefore the proposed framework tries to address issues related to both data sources and techniques applied to these data sources.

VI. Discussion

Many reviews and surveys have been done in the past to address the issue of drug safety and healthcare. In 2015, Lardon et al. [11] in their research study tried to explore the breadth of evidence about the use of social media as a new source of knowledge for pharmacovigilance. They adopted a similar methodology of collecting research articles based on research questions and then analyzing them from multiple perspectives. The scope of their work is satisfactory but they have limited themselves to only unstructured datasources and NLP (Natural Language Processing) tools and techniques while in our research study the author has provided a comprehensive approach in terms of dataset selection and tools and techniques applied to them. Another research study was done by Ho et al. [44] in 2016 collected and analyzed research papers in terms of their problem statement, the dataset used and the methodology applied. The research summary provided in this paper is sufficient but it does not lead to any concrete solution to the existing research problem. Similarly, research studies conducted by Tan et al. [45] in 2016 have reviewed the interaction of different ADR datasets with biological and genetic datasets. Further, they discussed the benefits and limitations of these integrated datasets in the current scope. The drug-ADR associations are analyzed statistically only on the basis datasource but no practical implementation is provided unlike in our research study. Although many other reviews and survey reports have discussed the major datasources related to ADR and their transition from a data-driven approach to machine learning models [42] for ADR prediction they do not provide an overall broad approach in terms of the datasources discussed, methodologies applied and a practical solution to the problem in the existing research works. Thus, our research study not only provides a comprehensive framework for both datasources and techniques applied to them but also implements the proposed model to obtain better results in terms of accuracy, F1 score, and interpretability.

VII. Conclusion

In conclusion, this research study provides a bird's eye view of drugs, the importance of drug-ADR association, and the methodologies used to discover them. It also analyses its impact on human health. Although each step in this research study has been carried out in detail starting from research paper selection to proposed framework implementation and results in discussion, still some research gaps in the given study that should be considered for future research. First, the research papers are selected based on single drug-ADR association while research studies considering drug-drug interactions are ignored, so for future research work research studies considering drug-drug interaction should also be included. The proposed model has applied only a deep neural network for prediction and evaluated its performance based on the different number of hidden layers. Further, the author proposes to apply different deep learning models to this integrated dataset and then compare its performance with the existing results. The main aim of this research is to optimize the performance of the proposed model in terms of accuracy and other evaluation metrics.

Acknowledgment

We would like to extend our sincere gratitude to Prof. Shreerang V. Joshi and Prof. Prashant S. Kharkar of the Institute of Chemical Technology (ICT), Mumbai as domain experts for their invaluable assistance and support in helping us to understand the pharmacological aspect of our research work. Their insights into our research have greatly improved the manuscript.

References

[1] DL. Patrick, et al., “Patient-reported outcomes to support medical product labeling claims: FDA perspective.” Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, vol. 10, Suppl 2, pp. S125-37, 2007.

[2] L. Hazell, Lorna, and SAW. Shakir, “Under-reporting of adverse drug reactions : a systematic review.” Drug safety vol. 29,no. 5, pp. 385-96, 2006.

[3] The ICH Expert Working Group. (Nov. 2003). Post-approval safety data management: Definitions and standards for expedited reporting.

[4] J.Lazarou, et al., “Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies.” JAMA vol. 279,no.15, pp.1200-5,1998.

[5] DK.Wysowski and L. Swartz, "Adverse Drug Event Surveillance and Drug Withdrawals in the United States, 1969–2002: The Importance of Reporting Suspected Reactions". Archives of Internal Medicine. Vol.165,no.12,pp.1363–9, 2005.

[6] "Sibutramine (brand name Reductil) Information – Australia". Abbott Laboratories. 2010. Archived from the original on 2010-10-14. Retrieved 2010-10-08.

[7] J.Jin, et al., "Factors affecting therapeutic compliance: A review from the patient's perspective." Therapeutics and clinical risk management vol. 4,no.1, pp. 269-86,2008.

[8] M. Alomar.,"Factors affecting the development of adverse drug reactions". Saudi Pharmaceutical Journal, Volume 22, Issue 2, Pages 83-94,2014.

[9] G.Simper, et al., "Physiology and Pathology of Drug Hypersensitivity: Role of Human Leukocyte Antigens". Physiology and Pathology of Immunology, edited by Nima Rezaei, IntechOpen, 2017.

[10] H.Arksey and L. O'Malley, "Scoping studies: towards a methodological framework." International journal of social research methodology vol. 8, no. 1, pp. 19-32, 2005.

[11] J.Lardon, R. Abdellaoui, et al., "Adverse drug reaction identification and extraction in social media: a scoping review." Journal of medical Internet research vol. 17, no. 7, pp. e4304,2015.

[12] R.Liu and P. Zhang, "Towards early detection of adverse drug reactions: combining pre-clinical drug structures and post-market safety reports." BMC medical informatics and decision making vol. 19, no. 1, pp. 1-9,2019.

[13] MS.Islam, et al., "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining." Healthcare (Basel, Switzerland) vol. 6,no.2, pp. 54,2018.

[14] R. Ietswaart, et al., "Machine learning guided association of adverse drug reactions with in vitro target-based pharmacology." EBioMedicine vol. 57, pp.102837, 2020.

[15] NIH(National Library of Medicine), https://www.nlm.nih.gov/bsd/medline.html.

[16] PubMed, https://pubmed.ncbi.nlm.nih.gov/about/.

[17] Office of Scholarly Communication (December 2016). "A social networking site is not an open access repository". University of California.

[18] ResearchGate, https://www.researchgate.net/topic/Publication/.

[19] MeSH (Medical Subject Headings), https://www.ncbi.nlm.nih.gov/mesh/.

[20] JC.Talbot, J C, and B S Nilsson, "Pharmacovigilance in the pharmaceutical industry." British journal of clinical pharmacology vol. 45,no.5, pp. 427-31,1998.

[21] P. Biswas, Pipasha, "Pharmacovigilance in Asia." Journal of Pharmacology and Pharmacotherapeutics vol. 4, no. 1_suppl, pp. S7-S19, 2013.

[22] D.Pappa and LK. Stergioulas, "Harnessing social media data for pharmacovigilance: a review of current state of the art, challenges and future directions." International Journal of Data Science and Analytics vol.8, no.2, 113-135,2019.

[23] JM. Banda, et al., "A curated and standardized adverse drug event resource to accelerate drug safety research." Scientific data vol.3, no. 1 pp. 1-11,2016.

[24] Dal Pan G. J., Arlett P. R., 'The US Food and Drug Administration-European Medicines Agency collaboration in pharmacovigilance: common objectives and common challenges', Drug Saf., Vol. 38, pp. 13-15.

[25] B.Hugman, "From the Uppsala monitoring centre." Drug safety 28, no. 7, pp. 645-646, 2005.

[26] K.Häyrinen, K. Saranto, and P. Nykänen, "Definition, structure, content, use and impacts of electronic health records: a review of the research literature." International journal of medical informatics vol. 77, no. 5,333 pp. 291-304,2008.

[27] M.Kuhn, et al., "The SIDER database of drugs and side effects." Nucleic acids research vol. 44, no.D1, pp. D1075-9,2016.

[28] NP. Tatonetti, et al., "Data-driven prediction of drug effects and interactions." *Science translational medicine* vol. 4, no. 125, pp. 125ra31-125ra31, 2012.

[29] MC Cai,Q. Xu, et al., "ADReCS: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms." Nucleic acids research 43, no. D1, pp.D907-D913,2015.3.

[30] DS. Wishart,et al., "DrugBank: a knowledgebase for drugs, drug actions and drug targets." Nucleic acids research vol.36, no. suppl_1, pp. D901-D906,2008.

[31] Y. Wang, et al., "PubChem: a public information system for analyzing bioactivities of small molecules." Nucleic acids research vol. 37,S no. suppl_2, W623-W633,2009.

[32] VB Siramshetty, OA Eckert, et al., "SuperDRUG2: a one stop resource for approved/marketed drugs." Nucleic acids research vol.346, no. D1,pp.D1137-D1143,2018.

[33] M. Kuhn, et al., "STITCH: interaction networks of chemicals and proteins." Nucleic acids research vol 36, no. suppl_1, pp.D684-D6883, 2007.

[34] CF. Thorn, TE. Klein, and RB. Altman, "PharmGKB: the pharmacogenomics knowledge base." In Pharmacogenomics, pp. 311-320. Humana Press, Totowa, NJ, 2013.

[35] M.Kanehisa and S. Goto,"Comprehensive gene and pathway analysis of cervical cancer progression." Nucleic Acids Res vol.28, pp. 327-30,2000.

[36] DP. Hill, et al., "Gene Ontology annotations: what they mean and where they come from." In BMC bioinformatics, vol. 9, no. 5, pp. 1-9. BioMed Central, 2008.

[37] A. Czarnecki and S. Voss, "Safety signals using proportional reporting ratios from company and regulatory authority databases." Drug Information Journal 42, no. 3, pp. 205-210,2008.

[38] M.Wang,X. Ma,et al. "Adverse drug reaction discovery using a tumor-biomarker knowledge graph." Frontiers in genetics vol.11, pp. 625659, 2021.

[39] NEH. Daoud, et al., "ADMET Profiling in Drug Discovery and Development: Perspectives of In Silico, In Vitro and Integrated Approaches." Current Drug Metabolism vol. 22, no. 7, pp. 503-522, 2021.

[40] Y. Xia, "Correlation and association analyses in microbiome study integrating multiomics in health and disease." Progress in Molecular Biology and Translational Science vol. 171, pp. 309-491, 2020.

[41] A. Creswell,et al., "Generative adversarial networks: An overview." IEEE signal processing magazine vol.35, no. 1, pp.53-65,2018.

[42] CY.Lee and YP. Yen, "Machine learning on adverse drug reactions for pharmacovigilance." Drug discovery today vol. 24,no.7,pp. 1332-1343,2019.

[43] R.ElShawi,Y. Sherif, M. Al-Mallah, and S. Sakr, "ILIME: local and global interpretable model-agnostic explainer of black-box decision." In European Conference on Advances in Databases and Information Systems, pp. 53-68. Springer, Cham, 2019.

[44] TB. Ho,L. Le, DT. Thai, and S.Taewijit, "Data-driven approach to detect and predict adverse drug reactions." Current pharmaceutical design vol. 22, no. 23, pp. 3498-3526, 2016.

[45] Y. Tan, Y. Hu, et al., "Improving drug safety: From adverse drug reaction knowledge discovery to clinical implementation." Methods vol. 110, pp.14-25, 2016.