iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XI **Month of publication:** November 2022

DOI: https://doi.org/10.22214/ijraset.2022.47504

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XI Nov 2022- Available at www.ijraset.com

Drug Discovery Using Machine Learning and Data Analysis

Pranali Pali[1], Abhishek Jain[2], Pooja Kyadarkunte[3], Sampada Patil[4], Shruti Jais[5], Rajesh Nasare[6], Sharda Chhabria[7]
[1, 2, 3, 4, 5]*Student,* [6, 7]*Mentor, Department Of Artificial Intelligence, G H Raisoni Institute Of Engineering And Technology Nagpur*

Abstract: ***An objective of drug discovery is to identify novel substances with certain chemical properties for the treatment of diseases. A significant amount of biological data has been produced recently from a variety of sources. Using this data, molecular analysis has been used to determine the most successful treatments. Trial-and-error medicine is frequently frustrating and significantly more expensive. This makes it easier to complete the work by predicting whether a drug will be active or not. The information about the drug can also be used to develop new medications. Quantitative Structure Activity Relationship (QSAR) analysis is one application that uses machine learning to improve decision-making in pharmaceutical data across multiple applications. Predictive models based on machine learning have recently grown substantially in prominence with in phase beyond preclinical research. In this stage, new drug discovery expenses and research times are significantly reduced. Utilizing pattern recognition algorithms, deciphering mathematical correlations, chemical and biological features of compounds, and machine learning has been used for drug development increasingly and more frequently, with positive outcomes. Other restrictions include the necessity for a large volume of data, a lack of interpretability, etc. Machine learning approaches are comparable to physical models in that they may be applied to large data sets without the need for computational resources.***

Index Terms: Quantitative Structure Activity Relationship (QSAR), Pharmaceutical Data, Preclinical Research, Pattern Recognition, Mathematical Correlations.

I. INTRODUCTION

Identifying a therapeutically effective molecule for the diagnosis and treatment of disease has been the purpose of the drug discovery process. According to the Precision Medicine Initiative, precision medicine is a new approach to disease treatment and prevention that considers a person's unique genetic makeup as well as their surroundings and way of life. With the use of this innovative method, medical professionals and researchers can now anticipate with greater precision which disease treatment and preventative techniques will be effective for which demographics. The search for innovative medications continues to be a time-consuming and expensive process. A new medicine typically takes between 10 and 15 years to produce through research and testing. The cost of drug discovery and development is dramatically rising when using a standard approach. There are many different chemical compounds with different properties that have existed for ages. Building a medication for the particular ailment will be made easier by studying these chemicals. Whole-person care is the foundation for drug research. By using conventional procedures, finding drugs takes a long time. This project uses machine learning approaches to address the aforementioned issues. This model's major purpose is to cut down on expenses and time spent on research when finding new drugs. The ChEMBL database, which contains information on compounds' biological and chemical properties, is used in this study. Acetylcholinesterase is particularly mentioned as a compound in this research.Let us understand some of the important terminologies :

1) *Drug Discovery:* The identification and confirmation of a disease target and the discovery and development of a chemical compound to interact with that target can be summed up as the drug discovery process. Drug development begins with the identification of an illness with well defined symptoms that lower quality of life. A desirable drug is typically defined as a chemical (which could be a simple molecule or a sophisticated protein) or chemical combination that lessens symptoms without having a significant negative impact on the patient.
2) *Bioinformatics:* The field of bioinformatics focuses on the exponential rise of biological data, which has sparked the creation of primary and secondary databases of protein and nucleic acid sequences and structures. It is a subdivision of science that studies biological data utilising computational methods as well as statistical tools, algorithms, and mathematical principles. It is believed that bioinformatics is important for comprehending the complicated cell mechanisms. Additionally, biomedical researchers find that bioinformatics is particularly helpful for analysing clinical samples.
3) *Artificial intelligence (AI):* Artificial intelligence is the study of the wide-ranging applications and organisational structures of various algorithms for deciphering and drawing knowledge from data. Additionally, the concept of AI is closely related to a wide range of fields, including pattern recognition, probability theory, statistics, machine learning, and a variety of techniques

including fuzzy models and neural networks, all of which fall under the umbrella term "Computational Intelligence". It has the capacity to address a variety of issues relating to human intelligence and, in turn, the simulation of these processes using computer systems or software. The science of artificial intelligence has now evolved from theoretical understanding to an actual data.

4) *Machine Learning(ML):* Machine learning is a branch of artificial intelligence that depends on computational and mathematical theory. Machine learning is based on creating models through training data exposure. Within pharmaceutical companies, the use of ML algorithms has significantly increased during the past two years. Instead of being limited to particular data kinds in the past, such as protein sequences and chemicals, it may be used with a wide range of data types and methodologies, including imaging and protein structures. The use of machine learning for drug development has been steadily expanding, and it is producing effective results by utilising pattern recognition algorithms, astute mathematical correlations, etc.
5) *Deep Learning:* A component of machine learning called "deep learning" can use multiple layers of input data to extract a higher degree of features. Deep learning is a vast field that currently commands enormous premiums. Deep learning algorithms have recently gained greater practicability in business strikes and have been employed in several scientific fields. How, though, does deep learning work? In general, deep learning uses the same neural network architecture, which has multiple layers and allows for data transformations between them. The innovation that led to its continued widespread use is real and well-considered. Therefore, models in deep learning can be created using a method known as greedy layer-by-layer.

II. METHODOLOGY

Even though all study domains share some steps in the experimental design, the use of an ML approach must be cross-disciplinary. We can distinguish the following steps in the ML methodology used in drug discovery specifically: Data Collection, Creating Mathematical Descriptors, Searching for best selection of variables, Model Training, and Model Validation are the essential five steps:

1) *Data Collection:* Getting the data set, which must meet certain requirements, is the initial stage. It must have properties that make it simple to generate and manage in the lab in addition to physical-chemical qualities that aid absorption, specificity, and low toxicity. This is because complicated compounds or big proteins are not commonly used in the pharmaceutical sector. Small molecules and peptides are the major types of chemicals that it typically interacts with. The sequencing and structure of small molecules and peptides, respectively, are represented using the SMILES(simplified molecular-input line-entry system) and FASTA formats to facilitate the handling and analysis of these substances. The field of drug discovery currently has a wide number of public repositories 1 that house valuable data, including DrugBank, PubChem , ChEMBL , and ZINC.

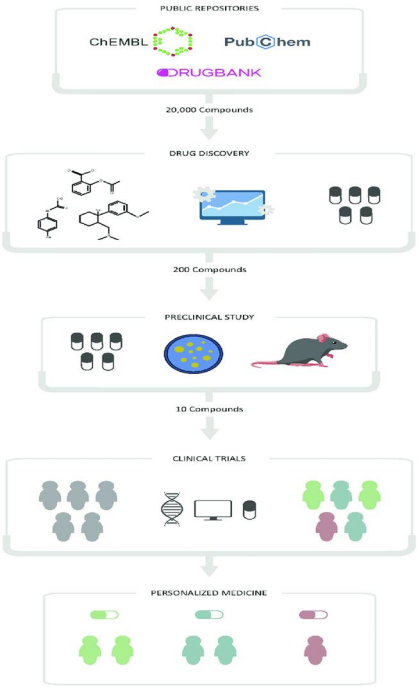


Fig.1 Stages of new drug discovery

2) *Creating Mathematical Descriptors:* Although certain machine learning (ML) models do not require labelling, supervised learning models are frequently used in the field of drug development. In this instance, the researchers' defined labelling will be crucial to the experimental procedure. A set of data that the ML model can process is obtained with the production of the mathematical descriptors. This dataset is split into two subsets: one with a larger percentage of data used to train the model (shown in Fig. 2.) and a smaller one used to test the model (represented in Fig. 2.). With the correct and required information, the best subset of variables inside the training set is sought after. Unsurprisingly, a lot of numerical variables are offered when creating mathematical descriptors. The fundamental goal of this approach is to eliminate as many redundant or superfluous variables as feasible. To this goal, other methods exist, including PCA(Principal Component Analyisis), t-SNE(t-Distributed Stochastic Neighbor Embedding), FS(Feature Selection), Autoencoder, etc.

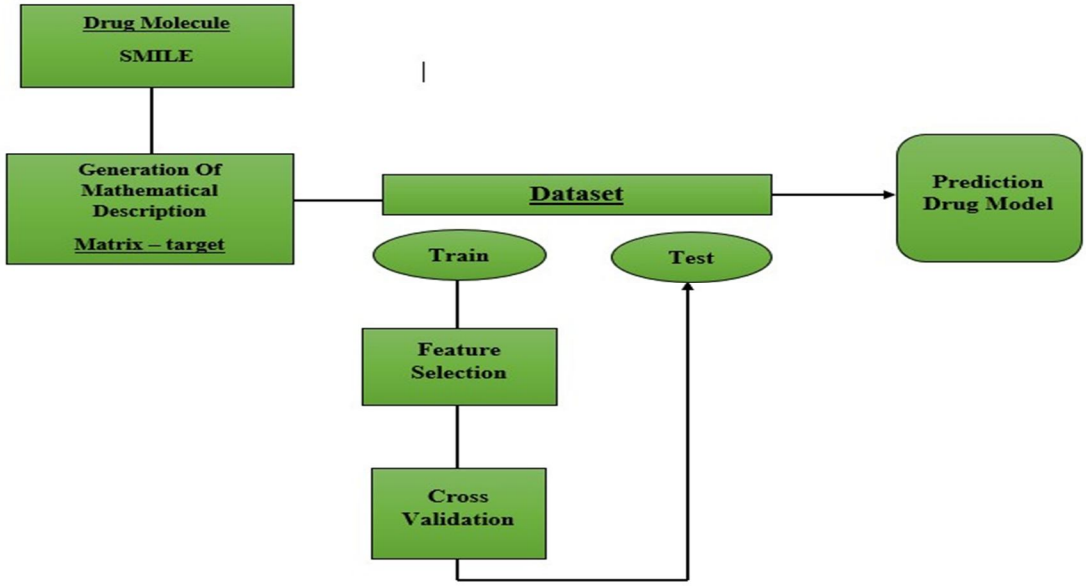


Fig.2 Machine Learning Methodology

3) *Searching for Best Selection of Variables:* A subset of the original collection of features is selected using FS approaches, but not alter the variables' contents. The algorithms and their input parameters must be chosen first. To make sure they are appropriate for the task at hand and the quantity and type of data available, these must be carefully picked. This provides a biologically comprehensible justification, which is the reason why a Most researchers employ these methods while creating their experimental designs.
4) *Model Training:* The model is trained when the best collection of variables has been identified. The experiment is then run a number of times using the practise data. To ensure the model's applicability to unknowable inputs, excessive training should be avoided. In these situations, cross-validation (CV) techniques are frequently used. The CV enables performance evaluation, performance estimation with unknown data, and monitoring the model's degree of generalisation throughout the training phase.
5) *Model Validation:* The initial data set is separated once more into three groups for each execution of the experiment. The training set and the validation set are two subsets. Figure 2 shows the evolution of the CV approach over the course of 10 runs. The blue set represents the training set for each of these runs, and the red set represents the validation set. The optimal parameter combinations for each approach are to be chosen as the final result of the CV process. Using these criteria, each model's performance is evaluated. The model with the highest performance value at the lowest total cost is the best one. The test set that was taken from the original set is then retrieved (shown in Fig. 2), and the best model that was produced by the CV process is then final validated. A new predictive drug model may have been developed if the validation results are statistically significant.

The application of machine learning techniques is widespread, and more articles have been published recently in particular. However, there aren't many machine learning publications on open access platforms that are concerned with medication development.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XI Nov 2022- Available at www.ijraset.com

III. DATABASES, SOFTWARES, PACKAGES and THEIR REPRESENTATION

A. *Databases*

1) *ChEMBL:* The information contained in the ChEMBL database was manually gathered from works of literature. The European Bioinformatics Institute and the European Molecular Biology Laboratory (EMBL) released this database in 2002. This database now contains more than 1.9 million chemical compounds and still adding, since the most recent update in 2018. More than 10,000 medicines and more than 12 000 targets are contained in these compounds according to ChEMBL. Since it is a live dataset, you may access it by integrating it into an API and getting your data from that.
2) *DrugBank:* One of the most well-known databases and a frequently used source for drug information is DrugBank. 2006 saw the initial publication of this database. It is a bioinformatics and cheminformatics database that includes thorough drug target and detailed drug data. DrugBank's DTI connections were initially gathered from books, journals, and other electronic databases. It also offers free downloads of all data.
3) *PubChem:* Information about chemical compounds and their corresponding biological actives is kept in PubChem. Substance, Compound, and BioAssay are the three sub-databases that make up this database. The main repository for storing chemical data supplied by individual data contributors is called substance. The distinct chemical structures that were taken from the Substance database and stored in the Compound database. The BioAssay database contains all biologically relevant information about this chemical substance data.

B. *Softwares*

In order to make model interpretation easier, a number of software tools have been developed in light of the current interest deep learning applications are receiving. Captum, an addition to the PyTorch deep learning and automatic differentiation package that offers support for the majority of the feature attribution strategies discussed in this paper, is a notable example. Alibi, another well-liked package, offers instance-specific justifications for individual models developed using the scikit-learn or TensorFlow libraries. Anchors, descriptive explanations, and counterfactual examples are a few of the explanation techniques used.

C. *Packages*

Based on the prior work, Sakakibara created a web service called Comprehensive Predictor of Interactions between Chemical Compounds and Target Proteins, which use SVM as the Drug Target Interaction(DTI) predictor. It appears that this server is no longer accessible.

In order to integrate chemoinformatics, bioinformatics, proteochemometrics, and chemogenomics for DTI prediction, Cao created the Python tool PyDPI based on Random Forest. The proposed approach uses prepared dictionaries for categorization and requires choosing chemical characteristics. This package can be used to build web-based servers and offers an interface for databases including PubChem, Drugbank, Uniprot, and the Kyoto Encyclopedia of Genes and Genomes (KEGG). The same team also developed PreDPI-Ki, a web-based service, in the same year. PreDPI-Ki is built on a random forest predictor and considers the binding affinities of DT pairs to better anticipate interactions.

D. *Representation*

1) *SMILE Code:* The best technique to represent a molecule is to text-encode its structural information. In this procedure, graphical structural data is transformed into text, which is then employed in the machine learning pipeline.The most widely used illustration is called SMILES (Simplified Molecular -Input Line Entry System). Once the conversion is complete, we can process the medicine and anticipate its properties, chemical interactions, and side effects using other algorithms like NPL.
2) *Molecular Fingerprint:* A drug can be represented in the machine learning input pipeline in a variety of ways, one of which is via its molecular fingerprint. Binary digits are the most typical form and can indicate whether a molecule has a specific substructure or not. It is clear that using a molecule as a vector to encode information is a procedure that cannot be undone. This operation results in the loss of information because it is impossible to recover the fingerprint from the molecule.
3) *FASTA Code:* Base pairs or amino acids are represented using single-letter codes in the text-based FASTA format, which is used to represent either nucleotide sequences or peptide sequences. A single line of description is followed by several lines of sequence information to make up a sequence in the FASTA format. The greater-than (">") character in the first column designates the description line as distinct from the sequence data. It is advised that all text lines be no more than 80 characters.

IV. ALGORITHMS

1) *Naïve Bayes:* The prediction of potential drug targets has been done using this approach in drug discovery. They specifically created a Bayesian model that incorporates many data sources, including data on known side effects or gene expression, and they achieved a model with 90% accuracy on more than 2,000 compounds. They also created the experimental validation of the screening procedure. They use data from ChEMBL to generate two descriptors, which they then use to validate their predictions using docking methods. They then predict compounds that are multi-target with an AUC of 80% for the therapy of HCV. They created a model for the prediction of ligand-target interactions with a 95% accuracy using interactions with four different classes of proteins (enzymes, ion channels, GPCRs, and nuclear receptors) retrieved from KEGG and DrugBank as well as random interactions from STITCH.
2) *Support Vector Machine(SVM):* Due to its capacity to discriminate between active and inactive molecules as well as its capacity to train the regression model, SVM plays a significant role in drug discovery. Regression models are crucial to determining the interaction between drugs and ligands. It is also capable of handling challenging issues that are complex, non-linear, high-dimensional, and noisy. Based on their classification in KEGG, they have been used to classify medicines with an accuracy value of 83.9%. a new framework that has an F1 value of 80% accuracy for predicting complex drug-target interaction networks from interaction matrices. By calculating several molecular descriptors and chemical indices from 25 ChEMBL datasets, it is also possible to predict the stability in human liver microsomes with values near to 70% accuracy in validation.
3) *Random Forest(RF):* A collection of decision trees is called Random Forest. We'll have a selection of choice trees in the random forest. If we need to categorise an object based on an attribute, every tree will cast a vote for a class, and the classification with the most votes will be chosen by the decision forest. A single decision tree typically does not produce high-performance results. To reduce high variance, the tree is typically pruned using cross validation or model complexity parameters. It has been demonstrated that RF models enhance the LBVS performance of specific Decision Trees. The QSAR data prediction is enhanced by the characteristics of Random Forest. High prediction accuracy and descriptor selection are built-in properties. Regardless of the challenge at hand, this is one of the most often used ML methods, and while it is impossible to single out a model as being the best for all problem types, RF is unquestionably among the best in terms of performance, speed, and generalizability. They were able to predict compound-protein interactions with an accuracy of more than 90% using 211,888 compound-protein interactions from BindingDB in an mRMR (max relevance and min redundancy) dimensionality reduction scheme and the descriptors produced with Open Babel and the enrichment scores of each protein from GO and KEGG. They determined that the most accurate predictors of synergy score were tree-based models based on gene expression and mutation data in cancer-related pathways. It is calculated using the molecular descriptors with RDKit, and a new tree-based model with AUC values higher than 90% has been proposed. This model uses the Relief algorithm for feature extraction and Graph Based Semi-supervised Learning as a classifier.
4) *Artificial Neural Network(ANN):* The input layer of an ANN refers to the neurons that receive data from the outside world and are referred to as input nodes. Additionally, the network requires output nodes, which are located in the hidden layer and transmit ANN results. The remaining nodes are categorised into one or more hidden layers and are known as hidden nodes, which transport information between neurons in the network. They employed 2D descriptors, which had information on the presence or absence of functional groups inside the molecule, and 1D descriptors, which contained information about the overall molecule (molecular weight, number of hydrogen bonds, etc.), for each of the compounds. An ANN with both types of descriptors produced the greatest results, with an accuracy of 89%. They forecast the early carcinogenesis of substances suggested to be medications for which they calculate six distinct types of descriptors with a deep learning model and an accuracy of 86% using 1003 chemicals from the Carinogenic Potency Database.
5) *K Nearest Neighbor:* It can be applied to classification and regression issues. It is primarily employed for classification issues. KNN classifies a new case using the majority votes of its k neighbours after storing all of the existing examples. To determine which of its k closest neighbours is the most prevalent, the distance function is used. These distance measures include the Makowski, Hamming, Manhattan, and Euclidean distances. It can be difficult to choose K when performing KNN modelling. A novel chemical is predicted using k neighbours, and this method is susceptible to noisy data. This data could result in misleading predictions in the molecule if the training data is incorrectly categorised.
6) *Decision Tree:* A supervised learning approach known as a decision tree is frequently employed for classification issues. Decision trees are used to classify the data and produce suggestions based on a set of decision rules. In the pharmaceutical industry, decision trees are used to solve issues like compound profiling, combinatorial library creation, drug likeness prediction, and more. Decision trees are also used to forecast the ADME features of p-glycoprotein, metabolic stability,

penetration, drug permeability, distribution, and solubility. Decision tree models are straightforward and simple to validate, comprehend, and understand. Decision tree predictions are known to have large variance. Splits in the results could emerge with even the smallest change in the data. The hierarchical structure is what leads to this instability.

7) *Lazy Regressor:* You can simplify your life with some packages. Lazy Regressor is one such software that ranks the machine learning models that are most likely to be appropriate. You can forecast binary and continuous variables, respectively, using the Lazy Predict's Lazy Classifier and Lazy Regressor. As it enables the training of 30 models at once and simplifies the training process, it is the optimal algorithm for effectively training your model.

V. CHALLENGES

1) The difficulties in producing accurate forecasts of Drug Target Identification can be divided into two groups: those involving databases and those involving computations. Depending on the nature of the task, one can frequently solve the computing challenges utilising various prediction methods. However, the sources of the databases presents significant challenges.
2) Research in pharmaceutical firms has generally relied on the integration of heterogeneous data, which presents its own obstacles in a variety of contexts and scales, from enormous molecules to specific persons. For handling multiple sources, a high level of artificial intelligence must be attained, and it must be enhanced with a greater comprehension of the data gathered. In order to consolidate the disparate data, contemporary data connectors are advised. Finally, these data connectors assist in allocating original data.
3) The problem of accessibility is one more difficulty in drug discovery because different classification models have ambiguous decision-making. Numerous mechanisms must be understood in medication development in order to evaluate the results. As a result, it is more helpful in identifying potential drug targets, and numerous assembled characteristics must increase interpretability confidence. Numerous methods, including SVM, Machine Learning, Random Forest, and deep learning algorithms, can be used in drug development to understand and interpret the results. Therefore, it is more helpful in identifying new drug targets and numerous assembled aspects for fostering interpretability.

VI. CONCLUSION

In the world of medicine, ML models can replace more traditional methods like PPT inhibitors and macrocycles by making predictions based on learned data inside a known framework, i.e., the compound structure. Deep learning models can also take into account chemical structures and QSAR models from pharmaceutical data because they were pertinent for molecules with the correct characteristics and had a high clinical trial success rate. Deep learning techniques and machine learning algorithms are frequently employed in the pharmaceutical business. In drug development and healthcare service hubs, notably in image analysis and omics data, several problems have been overcome using ML algorithms. AI technology has improved by entering computer-aided drug development in an effort to regain the powerful skills in data mining. The development of machine learning methodologies and disciplines will benefit from the proliferation of data. The application of these models in cheminformatics, and more specifically in drug development, has greatly benefited the pharmaceutical industry. The use of descriptors derived from the structure of peptides or small molecules was the sole tool accessible up until this time. ANN have been used more recently to directly recreate graph-based molecules. As the area develops, researchers are looking for new drugs, treatments, or cures that are more efficient than those that are already accessible. Understanding the fundamental processes behind disease progression, the effects of already available medications, and the genetic makeup of patients can aid in the development of novel, highly targeted therapeutic therapies that will eventually improve patients' health and quality of life.

VII. FUTURE SCOPE

Future study should concentrate on techniques that incorporate various similarities. In comparison to methods that just employ one form of similarity, ensemble-based models are more likely to produce accurate findings. Repurposed medications, for instance, have been discovered by accident, pharmacological analysis, or retrospective clinical analysis (such as examining adverse effects). Research is now concentrating on the most effective ways to adopt a more comprehensive, systemic approach in light of the surprisingly successful early examples (repurposing minoxidil from hypertension to hair loss, sildenafil from angina to erectile dysfunction, and thalidomide from morning sickness to multiple myeloma). Medical science and web innovation have been combined to increase the predictive power of deep learning algorithms about biomarkers, side effects of treatments, and therapeutic benefits. Success in clinical trials is attained by the use of certain applications. Therefore, motivation for potential investments in pharmaceutical firms is carried out.

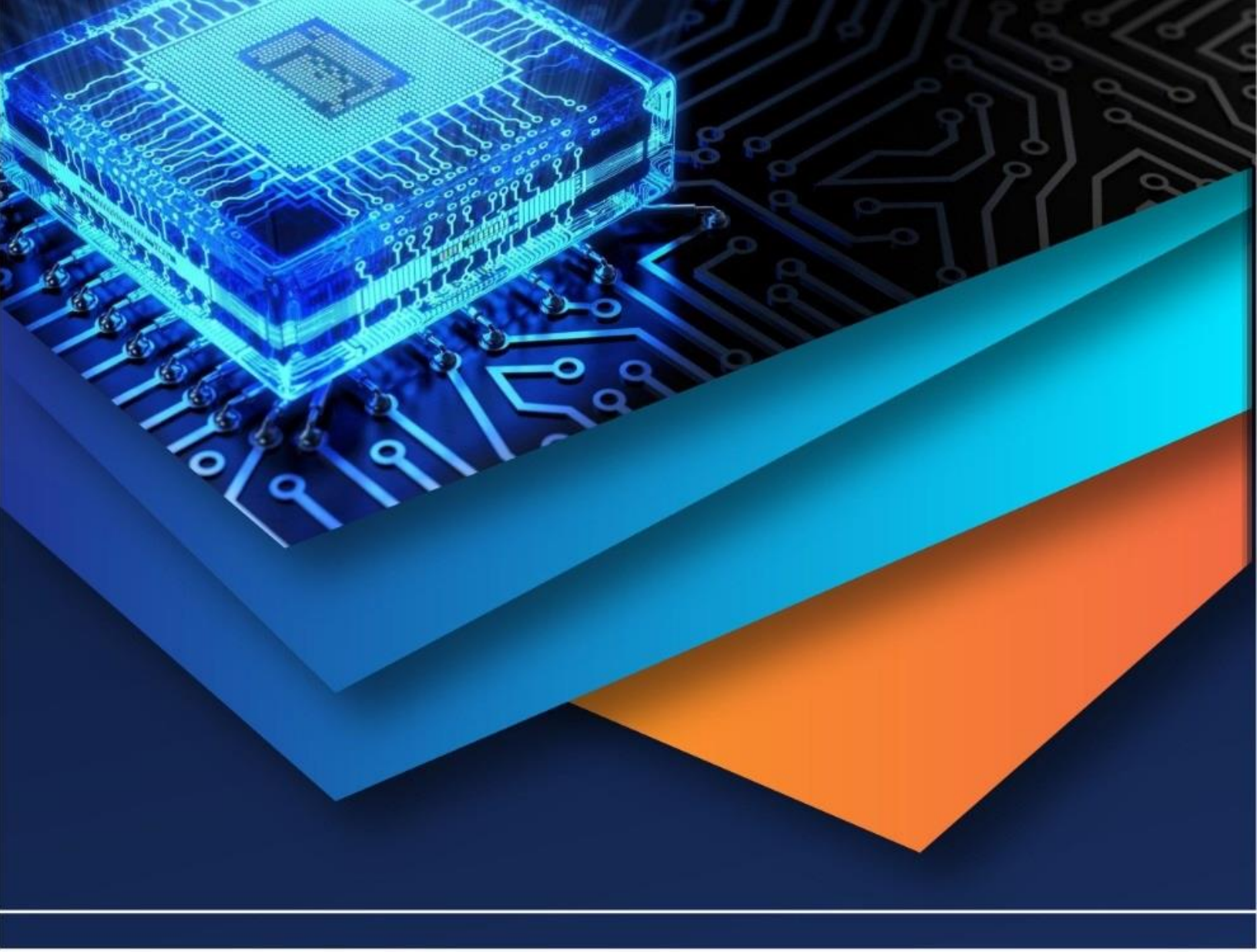
Future medication discovery and development plans anticipate using AI technologies to address every element. For new applications, automated AI has to coordinate theoretical findings like chemistry data, omics data, and medical data. Additionally, we anticipate that more confirmations will need to be rebuilt for the drug reveal campaign.

VIII. ACKNOWLEDGEMENT

The author would like to express their appreciation to Prof. Rajesh Nasare for his wise counsel and ongoing assistance during the project. They would also want to give particular thanks to Dr. Sharda Chhabria for her diligent supervision of the improvisation. We successfully finished this paper with your helpful guidance, and we are grateful to have both of you as our mentors.

REFERENCES

[1] Speck-Planche, M.N. Cordeiro, "Computer-aided drug design,synthesis and evaluation of new anti-cancer drugs", Curr Top MedChem. [Epub ahead of print], 2013.
[2] "Drug discovery hit to lead", Available from in Wikipedia.org/wiki/Drug_discovery_hit_to_lead, cited march 26 2012.
[3] Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. Nature 577, 706–710 (2020).
[4] Schneider, P. et al. Rethinking drug design in the artifcial intelligence era.Nat. Rev. Drug Discov. 19, 353–364 (2020).
[5] Karpov, P., Godin, G. & Tetko, I. V. Transformer-CNN: Swiss knife for QSAR modeling and interpretation. J. Cheminform. 12, 17 (2020).
[6] Zhang, R., Li, C., Zhang, J., Chen, C. & Wilson, A. G. Cyclical stochastic gradient MCMC for Bayesian deep learning. Preprint at https://arxiv.org/abs/1902.03932 (2019).
[7] Scalia, G., Grambow, C. A., Pernici, B., Li, Y.-P. & Green, W. H. Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. J. Chem. Inf. Model. 60, 2697–2717 (2020)
[8] Rifaioglu AS, Atas H, Martin MJ, et al. Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. Brief Bioinform,2018.
[9] Patel L, Shukla T, Huang X, Ussery DW, Wang S. Machine Learning Methods in Drug Discovery. Molecules 2020.
[10] Lavecchia A. Machine-learning approaches in drug discovery: methods and applications. Drug discovery today 2015.
[11] Tong WD, et al. Decision forest: combining the predictions of multiple independent decision tree models. J. Chem. Inf. Comput. Sci 2003.
[12] Feng Q, Dueva E, Cherkasov A, Ester M. Padme: A deep learning-based framework for drug-target interaction prediction. arXiv preprint arXiv:1807.09741 2018.
[13] https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/
[14] L. Burggraaff, P. Oranje, R. Gouka, P. van der Pijl, M. Geldof, H.W. van Vlijmen, A.P. IJzerman, G.J. van Westen, Identification of novel small molecule inhibitors for solute carrier sglt1 using proteochemometric modeling, Journal of cheminformatics 11 (1) (2019).
[15] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D.Johnson, C. Li, Z. Sayeeda, et al., Drugbank 5.0: a major update to the drugbank database for 2018, Nucleic acids research 46 (D1) (2018).
[16] A. Alimadadi, S. Aryal, I. Manandhar, P.B. Munroe, B. Joe, X. Cheng, Artificial intelligence and machine learning to fight covid-19 (2020).
[17] J.B. Cross, Methods for virtual screening of gpcr targets: Approaches and challenges, in: Computational Methods for GPCR Drug Discovery, Springer,2018.
[18] K. Zhao, H.-C. So, Using drug expression profiles and machine learning approach for drug repurposing, in: Computational methods for drug repurposing, Springer, 2019.
[19] Nicolas, J., Artificial intelligence and bioinformatics, in A Guided Tour of Artificial Intelligence Research. 2020.
[20] Kohli, A., et al., Concepts in US Food and Drug Administration regulation of artificial intelligence for medical imaging. American Journal of Roentgenology, 2019.
[21] Zhou, Y., et al., Artificial intelligence in COVID-19 drug repurposing. The Lancet Digital Health, 2020.
[22] Jiménez-Luna, J., F. Grisoni, and G. Schneider, Drug discovery with explainable artificial intelligence. Nature Machine Intelligence, 2020.
[23] Mohanty, S., et al., Application of Artificial Intelligence in COVID-19 drug repurposing. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2020.
[24] Paul, D., et al., Artificial intelligence in drug discovery and development. Drug Discovery Today, 2020.
[25] Cano G, Garcia-Rodriguez J, Garcia-Garcia A, Perez-Sanchez H, Benediktsson JA, Thapa A, Barr A Automatic selection of molecular descriptors using random forest: Application to drug discovery (2017).
[26] Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, Zhou J, Wang G Low-dose ct with a residual encoder-decoder convolutional neural network (2017).
[27] Mei J-P, Kwoh C-K, Yang P, et al. Drug–target interaction pre-diction by learning from local information and neighbors. Bioinformatics 2012.
[28] You J, McLeod RD, Hu P. Predicting drug–target interac-tion network using deep learning model. Comput Biol Chem 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY