REVIEW ARTICLE

UPI JOURNAL OF CHEMICAL AND LIFE SCIENCES (UPI-JCLS)

ISSN: 2581-4648 (An International online Peer Reviewed Open Access Journal)

www.uniquepubinternational.com Published by Unique Pub International (UPI)

A REVIEW ON MACHINE LEARNING IN PHARMACEUTICAL APPLICATIONS

Burle Sairaj*, Konkimudusu Mouli, Chandu Babu Rao

Priyadarshini Institute of Pharmaceutical Education and Research 5th Mile, Pulladigunta, Guntur-522017. Andhra Pradesh, India.

***Corresponding Author**
Burle Sairaj

Received: 15 June 2025 Revised: 04 JULY 2025 Accepted: 22 JULY 2025

Abstract

Artificial intelligence (AI) and machine learning (ML) have revolutionized pharmaceutical research and development (R&D), offering transformative applications in drug discovery, formulation development, and clinical trials. These technologies integrate seamlessly within the Quality by Design (QBD) framework and Process Analytical Technology (PAT), enabling data-driven decision-making and process optimization. ML algorithms, including artificial neural networks (ANNs), deep learning models, and light gradient boosting machine (LGBM) algorithms, enhance the efficiency and accuracy of pharmaceutical formulation processes. Over the past two decades, AI/ML approaches have been increasingly utilized in drug discovery, aiding in lead optimization, target identification, and virtual screening. High-throughput screening and computational modeling have strengthened the predictive capabilities of ML techniques, improving drug candidate selection. Additionally, AI is now impacting clinical trial design, execution, and data analysis, further accelerated by the COVID-19 pandemic and increased reliance on digital technologies.

Despite these advancements, challenges remain, including data quality, regulatory considerations, ethical concerns, and the need for transparency and accountability. Regulatory bodies are developing frameworks to ensure the safety and efficacy of AI-driven drug development. Future advancements include multi-task learning, personalized medicine, and AI integration with robotics and automation, which promise to further streamline drug development. This review provides a balanced perspective on AI/ML in pharmaceuticals, discussing key concepts, case studies, and emerging trends.

Keywords:Machine Learning, Artificial Intelligence, Drug Discovery, Quality by Design, Deep Learning, Clinical Trials, Regulatory Guidelines

INTRODUCTION

There is a growing demand for the swift development of pharmaceutical products, which can significantly benefit from advanced computational techniques. Similar to other scientific domains, artificial intelligence (AI), particularly machine learning (ML) algorithms, has demonstrated immense potential in uncovering intricate relationships within multivariable data generated during pharmaceutical development. Both formulation composition and processing parameters can be effectively optimized while minimizing variability in the final product's quality.

Artificial intelligence (AI) and machine learning (ML) have experienced significant growth over the past decade, propelled by groundbreaking advancements in computational technology. These developments have notably enhanced our capacity to gather and analyze extensive datasets. Concurrently, the expenses associated with bringing new pharmaceuticals to market have escalated to prohibitive levels. In this paper, we use the term "R&D" to broadly encompass the research, scientific endeavors, and processes involved in drug development.

Machine learning can be used in drug discovery to analyses massive volumes of data and find prospective new drug candidates that might be successful in treating particular conditions [1]. For instance, high-dimensional genomics data are analyzed using deep learning algorithms, a form of machine learning algorithm that can analyses complex data sets and find new drug targets. Other methods include reinforcement learning, which involves optimizing the drug development process by training algorithms, and transfer learning, which involves training machine learning algorithms on data from related domains.

While machine learning has the potential to completely change the drug development process by allowing researchers too quickly and accurately find novel drug ideas by analyzing massive amounts of data [2]. While employing machine

learning for drug discovery has several drawbacks and limits, there are also a number of potential strategies that are being developed to get over these obstacles.

Over Of the Available AI Methodology

Concepts of artificial intelligence and machine learning have been introduced in the middle of the 20th century. The versatility of their application in various fields, including healthcare, development of new medicines, and (bio)medicine in general, has been growing ever since ML can be based on supervised and unsupervised learning algorithms.

There are many different types of ANNs. Multi-Layered Perceptron (MLP) is the most often used, as one of the simplest yet powerful networks [3].

The role of machine learning in analyzing vast amounts of data for drug discovery

In order to find new medication candidates, massive and complex data sets must be analyzed. Researchers can quickly and accurately analyze these enormous volumes of data with the aid of machine learning (ML. In biological and chemical data, machine learning algorithms can find patterns and associations that can be exploited to create novel medications.

The capacity of machine learning to analyze data in real-time is another benefit for drug research. Researchers can swiftly find new patterns and associations that can be exploited to generate novel medications by utilizing ML algorithms to analyze streaming data [4]. However, there are a number of difficulties and restrictions with using ML in drug development.

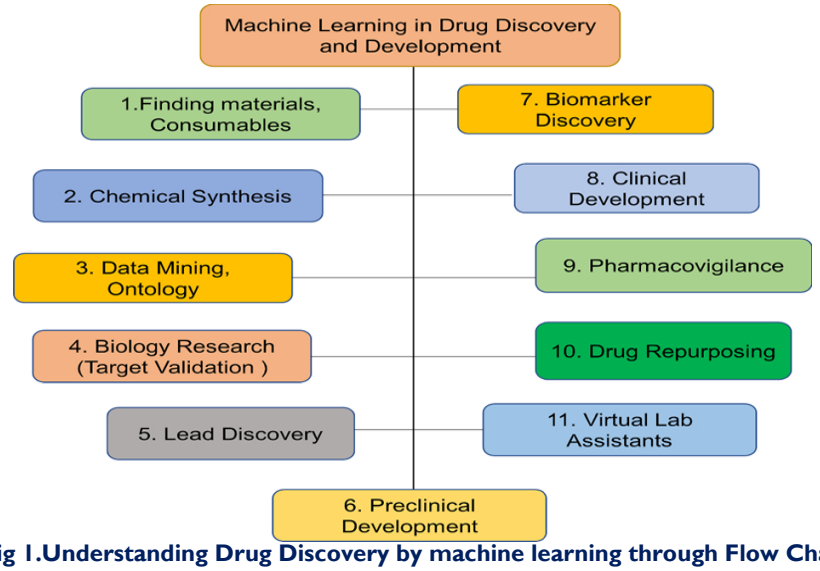

Fig 1.Understanding Drug Discovery by machine learning through Flow Chat

ML Procedures Used in Drug Discovery

ML Algorithms Used in Drug Discovery ML algorithms have significantly advanced drug discovery. Pharmaceutical companies have greatly benefited from the utilization of various ML algorithms in drug discovery. ML algorithms have been used to develop various models for predicting chemical, biological, and physical characteristics of compounds in drug discovery.

MLalgorithms and techniques are not a monolithic, homogeneous subset of AI. There are two main types of ML algorithms: Supervised and unsupervised learning. Supervised learning learns from training samples with known labels to determine labels of new samples. Unsupervised learning recognizes patterns in a set of samples, usually without labels for the samples [5].

Deep learning (DL) methods

DL algorithms are considered one of the cutting-edge areas of development and study in almost all scientific and technological fields. The renaissance of artificial NNs into workable algorithms from their former theorized and predicted applications, first developed in the 1950s, is an essential pillar of DL and the continued success brought by AI-based integration of standard techniques.

QSAR analysis is one of the most advanced forms of DL-based AI in current drug discovery and development. It has allowed researchers to take 2D chemical structures and determine physicochemical descriptors related to the molecule's activity. 3D-QSAR has allowed further inquiry of geometric structure impacting ligand-target interactions[6].

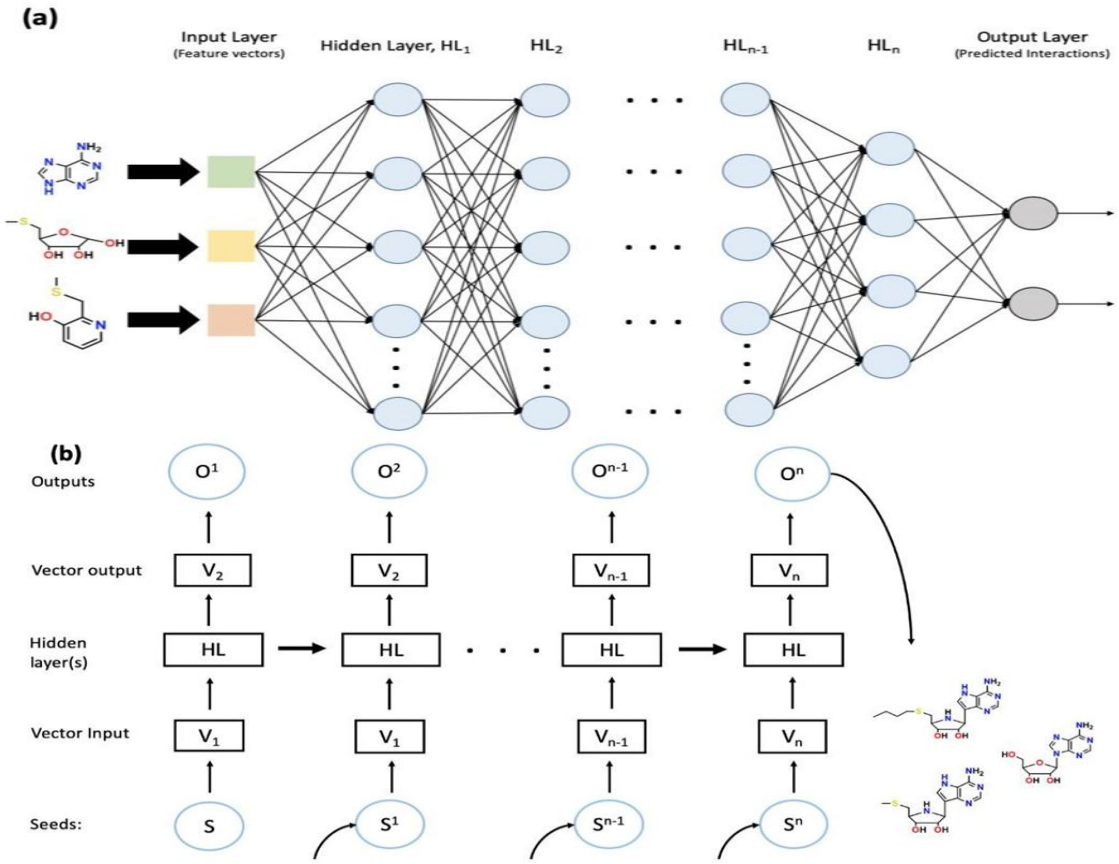

Figure-3: The general scheme of deep neutral network (DNN)

Challenges and limitations of using machine learning in drug discovery

The availability and quality of data are two of the main obstacles to applying machine learning in drug discovery. An extensive dataset of compounds with proven activity against certain biological targets is needed to train machine learning models to predict the biological activity of small molecules. However, such information is frequently sparse and of varying quality. As a result, the model's accuracy may suffer and biases may be introduced. Additionally, the data used to train machine learning models is frequently gathered from several sources and can be extremely varied in terms of the quality of the experimentation and the data itself.

The accuracy of the training data might be lower than anticipated. Even though algorithms discussed in this review have a higher threshold for minimizing errors, there are still some categorical errors from training sets[7].Amore concise way to understand this is by the statistical angle. With algorithms prediction, there is always a concern with overfitting or underfitting. Overfitting is when the model consists of lower quality information/technique but generates higher quality performance. It occurs when the model picks up unusual features during the training, resulting in a negative impact on the model. In contrast, underfitting models fail to recognize the data sets' underlying trend and generalize the new data inputted [8].

Deep learning in drug discovery

Using deep neural networks, quantitative structure-activity relationship (QSAR) modelling uses chemical structure to predict a compound's action against a particular biological target. To create a model that can be used to predict the activity of new compounds, the neural network is trained on a huge dataset of known compounds and their corresponding biological activities The action of hundreds of tiny compounds across a variety of biological targets has been predicted using this method. In one study, for instance, scientists utilized a deep learning model to forecast the activity of substances against the protein kinase CHK1, a target for cancer treatment. The model was developed by the researchers using a sizable dataset of well-known CHK1 inhibitors, and it was then used to forecast the activity of a group of novel drugs [9].

These models can be used to create novel compounds that are anticipated to have the desired activity after being trained. For instance, researchers created novel chemicals that were active against the protein target bromodomain-containing protein 4 (BRD4), a prospective target for the treatment of cancer, using a generative model. The model was trained using a sizable dataset of well-known BRD4 inhibitors, and it was then utilised to produce a collection of novel compounds. After testing these substances in vitro, the researchers discovered that several of them had a lot of action against BRD4. Deep learning can be used to forecast drug toxicity and side effects in addition to finding novel treatment candidates. This may lessen the number of medications that toxicity problems cause to fail in clinical trials. In order to forecast the toxicity of substances based on their chemical structures, for instance, researchers utilized a deep learning model.

The program was able to accurately estimate the toxicity of novel compounds after being trained on a sizable dataset of substances with known toxicity profiles. Deep learning, a subset of machine learning, has revolutionized the field of drug discovery by enabling the analysis of large datasets and identification of complex patterns. Convolutional Neural Networks (CNNs): CNNs are a type of deep learning algorithm that can be applied to drug discovery tasks such as molecular design, target prediction, and bioactivity prediction. Recurrent Neural Networks (RNNs): RNNs are another type of deep learning algorithm that can be used for tasks such as protein sequence analysis and molecular property prediction. Generative Models: Generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), can be used to generate novel molecular structures with desired properties. Target Prediction: Deep learning can be used to predict the targets of small molecules, enabling the identification of potential therapeutic applications. ADME Prediction: Deep learning can be used to predict the absorption, distribution, metabolism, and excretion (ADME) properties of small molecules. Toxicity Prediction: Deep learning can be used to predict the toxicity of small molecules, enabling the identification of potential safety issues[10].

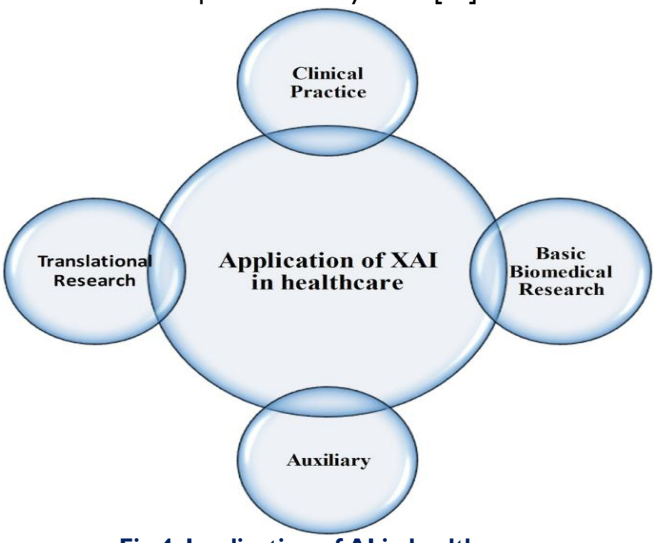

Fig.4. Implication of AI in health care

Application

Now that we have formalized the distinctive approaches used in ML and PMX, we review the trends, applications, limitations and promise of McLin pharmaceutical sciences.

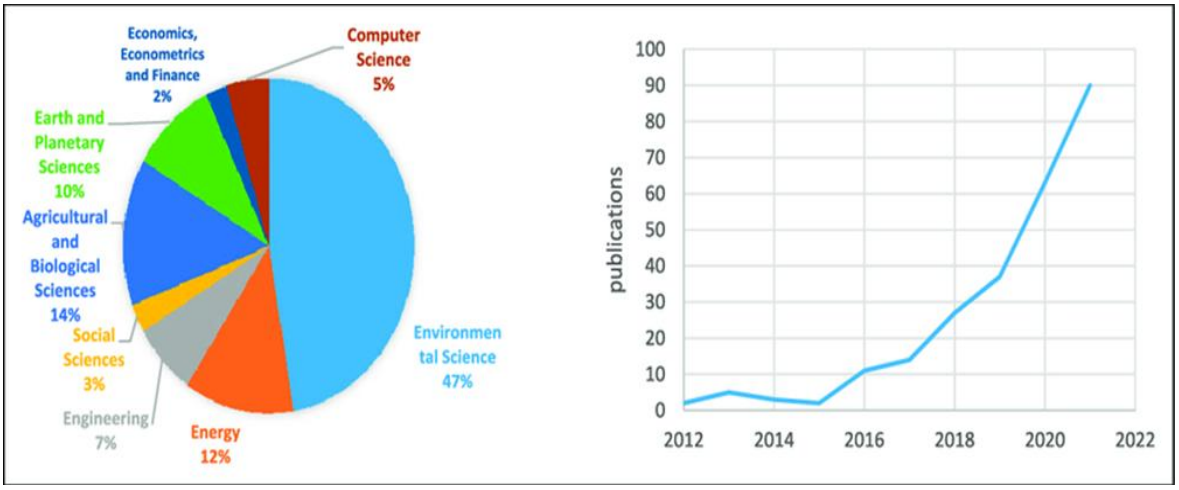

Fig 5: Publications of machine learning applications in pharmaceutical science2012-2022

As of July 2020, a Web of Science search of "machine learning" nested within the search of "pharmacokinetics or pharmacodynamics" yielded over 100 publications. The publications were categorized to 1 of the following areas of the pharmaceutical sciences: PK/PD, dose optimization, quantitative structure–activity relationship (QSAR); adverse drug event (ADE) prediction and drug drug interactions (DDIs) and clinical trial simulation (CTS). Figure summarizes the cat erotization of 65 ML publications in pharmaceutical sciences from 1995–2020. Supervised classification, regression, or a combination of the 2 approaches are the most frequently used ML methods and QSAR emerged as the most frequent application area. Table S1 sum maize's the pharmaceutical sciences applications and ML algorithms that have been investigated in the literature. As noted, previously several recent reviews have focused on identifying the most promising and appropriate application areas and ML tools. Talevi. provided an accessible primer on ML concepts that

included a case study that employed RFR for investigating the structure activity relationships of inhibitors of the putrescine transporter of trypanosome parasites.27,80 Hutchinsonetal. proposed an implementation framework with 2 hypothetical examples that utilized deep learning to perform global parameter sensitivity analysis and for combining PK parameters with imaging and omics data.25 Koch et al. used CART-based ML approaches for covariate selection in the context of a simulated data and clinically relevant example of phototherapy for bilirubinaemia in neonates.81 In a commentary, Chaturvedula et al. highlighted the use of genetic algorithms (GAs) for model selection in population modelling and deep learning for target identification in drug repurposing[11].

Drug Discovery: Machine learning (ML) is revolutionizing the drug discovery process by analyzing large datasets to identify potential therapeutic targets, predict drug efficacy and safety, and design new molecular structures. Target Identification: ML algorithms can analyze genomic data, protein structures, and disease mechanisms to identify potential therapeutic targets, reducing the time and cost associated with experimental screening. Lead Compound Identification: ML models can predict the efficacy and safety of potential lead compounds, enabling researchers to prioritize the most promising candidates and reduce the risk of costly failures.

ML can identify biomarkers for disease diagnosis, prognosis, and treatment monitoring, enabling personalized medicine and improving treatment outcomes Quality Control ML can detect anomalies and predict quality issues in real-time, ensuring consistent product quality and reducing the risk of product recalls. Supply Chain Optimization: ML can optimize supply chain operations, including inventory management, logistics, and distribution, reducing costs and improving efficiency.

Future Expectations

Serialization (track and trace) could also rely on ML tools. Although it is not its primary purpose, the data generated by serialization could also be used in product development and for tracking patients' adherence. Also, ML algorithms could be used for the identification of falsified medicines. More efficient manufacturing in the pharmaceutical industry can be expected with the integration of ML and PAT tools

Several such examples have been described in Table II. One of the greatest challenges related to successful PAT implementation is the analysis of high volume, multivariate data. Moreover, fast computations and decision making are of the utmost importance. These issues could be, potentially, solved by different ML tools. For example, Wongetal . have developed a method based on recurrent neural networks that provides efficient regulation of critical quality attributes. With many opportunities for the application of AI tools also come some obstacles and challenges, especially if algorithms and models are meant to be used in the mass production of medicines, either for pharmaceutical development or production process monitoring, or both. The challenges are related to the volume of data and speed of its accumulation; size of datasets; training learning time, over or under-fitting of models, etc. With the evolution of the big data concept and advances in computing capability, it is to be expected that the technical challenges will be reduced, but the necessity for critical considerations of AI-based models will still remain, as for any other modeling approach. In addition to these technical developments, the application of machine learning to drug discovery presents prospects for innovation.

The application of machine learning to the development of fresh medication candidates is also gaining popularity. In this method, referred to as generative modeling, new molecules with desired features are created using machine learning. By enabling the quick development and testing of numerous novel drug candidates, generative modelling has the potential to drastically cut down on the time and cost involved in drug discovery. Finally, there is a lot of room for machine learning to be used with other technologies, such automation and robotics, to speed up the drug development process even further [12].

Conclusion

Machine learning (ML) techniques are revolutionizing drug development by enhancing target discovery, lead compound discovery, synthesis, and protein-ligand interactions. ML applications in drug development rely on algorithm-enhanced data query, analysis, and generation. Target discovery is improved through ML-based refinement and search of existing omics and medical data. Viable targets can be identified using data clustering, regression, and classification from vast omics databases. Lead compound discovery is expedited using Quantitative Structure-Activity Relationship (QSAR) models. ML-based retrosynthesis algorithms accelerate lead compound synthesis with high accuracy. Decades of research have refined ML and deep learning techniques, enabling their successful application in drug discovery. Advances in computing power, algorithms, and investment have made ML applications more intelligent, cost-effective, and time-efficient. ML models have shown promise in developing new drug candidates, identifying novel therapeutic targets, and predicting pharmacological attributes. Future research directions include integrating ML with other technologies, addressing data quality and sharing challenges, and ensuring accountability and transparency in ML model application Machine learning is transforming the drug discovery process by enabling the analysis of vast amounts of data. This allows researchers to identify patterns and relationships that may not have been apparent through traditional methods. ML algorithms can be trained on large datasets to predict the efficacy and safety of potential drug candidates. This enables researchers to prioritize the most promising candidates and reduce the risk of costly failures.

Deep learning models, in particular, have shown great promise in drug discovery due to their ability to learn complex patterns in data. Multi-task learning is another area of research that holds great potential for drug discovery. This

approach enables ML models to learn from multiple related tasks simultaneously, improving their overall performance. Personalized medicine is another area where ML is having a significant impact, enabling researchers to tailor treatments to individual patients. The integration of ML with other technologies, such as robotics and automation, is also expected to accelerate drug discovery. However, to fully realize the potential of ML in drug discovery, researchers must address challenges related to data quality, sharing, and interpretation.

Author Contributions
All authors are contributed equally

Financial Support
None

Declaration of Competing Interest
The Authors have no Conflicts of Interest to Declare.

Acknowledgements
None

References
1. Harnie D, Saey M, Vapirev AE, Wegner JK, Gedich A, Steijaert M, Ceulemans H, Wuyts R, De Meuter W. Scaling machine learning for target prediction in drug discovery using Apache Spark. Future Generation Computer Systems. 2017 Feb 1;67:409-17. https://doi.org/10.1016/j.future.2016.04.023
2. Petrey D, Honig B. Structural bioinformatics of the interactome. Annual review of biophysics. 2014 May 6;43(1):193-210. https://doi.org/10.1146/annurev-biophys-051013-022726
3. Djuris J, Vidovic B, Ibric S. Release modeling of nanoencapsulated food ingredients by artificial intelligence algorithms. In release and bioavailability of nanoencapsulated food ingredients 2020 Jan 1 (pp. 311-347). Academic Press. https://doi.org/10.1016/B978-0-12-815665-0.00009-6
4. Deshpande M, Kuramochi M, Wale N, Karypis G. Frequent substructure-based approaches for classifying chemical compounds. IEEE Transactions on Knowledge and Data Engineering. 2005 Jun 27;17(8):1036-50. https://ieeexplore.ieee.org/abstract/document/1458698
5. Rifaioglu AS, Atas H, Martin MJ, Cetin-Atalay R, Atalay V, Doğan T. Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. Briefings in bioinformatics. 2019 Sep;20(5):1878-912. https://doi.org/10.1093/bib/bby061
6. Chen R, Liu X, Jin S, Lin J, Liu J. Machine learning for drug-target interaction prediction. Molecules. 2018 Aug 31;23(9):2208. https://doi.org/10.3390/molecules23092208
7. Kaiser TM, Burger PB. Error tolerance of machine learning algorithms across contemporary biological targets. Molecules. 2019 Jun 4;24(11):2115. https://doi.org/10.3390/molecules24112115
8. Lever J, Krzywinski M, Altman N. Points of significance: model selection and overfitting. Nature methods. 2016 Sep 1;13(9):703-5. https://doi.org/10.3390/molecules24112115
9. Rangwala H, Karypis G. fRMSDPred: Predicting local RMSD between structural fragments using sequence information. Proteins: Structure, Function, and Bioinformatics. 2008 Aug 15;72(3):1005-18. https://doi.org/10.1002/prot.21998
10. Chaitanya, GMSK; Sasi, B; Kumar, Anish; Rao, C Babu; Rao, B Purnachandra; Jayakumar, T; Prediction of fracture profile using digital image correlation Twelfth International Conference on Quality Control by Artificial Vision 2015, 9534-281-288-2015, SPIE https://doi.org/10.1117/12.2182911
11. Chaturvedula A, Calad-Thomson S, Liu C, Sale M, Gattu N, Goyal N. Artificial intelligence and pharmacometrics: time to embrace, capitalize, and advance?. CPT: pharmacometrics & systems pharmacology. 2019 Jun 5;8(7):440. https://doi.org/10.1002/psp4.12418
12. Salim N, Holliday J, Willett P. Combination of fingerprint-based similarity coefficients using data fusion. Journal of chemical information and computer sciences. 2003 Mar 24;43(2):435-42. https://doi.org/10.1021/ci025596j