

Role of artificial intelligence and machine learning in bioinformatics: Drug discovery and drug repurposing

Sameer Quazi,

Founder and CEO, GenLab Biosolutions Private Limited, Bangalore, Karnataka, India.

Abstract:

Artificial intelligence AI or machine learning has proven to be a potential activity in the health and biomedical sciences. Previous research it has found that AI can learn new data and transform it into the useful knowledge. In the field of pharmacology, the aim is to design more efficient and novel vaccines using this method which are also cost effective. The underlying fact is to predict the molecular mechanism and structure for increased likelihood of developing new drugs. Clinical, electronic and high-resolution imaging datasets can be used as inputs to aid the drug development niche. Moreover, the use of comprehensive target activity has been performed for repurposing a drug molecule by extending target profiles of drugs which also include off targets with therapeutic potential providing a new indication.

Key words:

Drug discovery, drug repurposing, bioinformatics, machine learning, artificial intelligence, biomedical discoveries etc.

1) Introduction:

Over the last few years, an immense progress has made in the fields of artificial intelligence, machine learning and bioinformatics and more research is needed for understand the data in biological science and related problems. Bioinformatics is a subdivision of science that involves the analysis of biological data using mathematical principles, statistical tools and certain algorithms in addition to computational approaches [1]. Artificial intelligence is the ability to solve various problems related to human intelligence and in turn the simulation of these intelligence processes using computer systems or software. It involves machine learning that allows one to perform all of these tasks based on its training [2]. Basic or structural bioinformatics tools use artificial intelligence and machine intelligence for the design of drugs and repurposing various novel compounds against many diseases such as cancer, neural inflammation and others using the silica approach by applying certain tools with principles of artificial intelligence and machine learning. Bioinformatics has been used to analyze data and logical conclusions. The huge amount of data obtained from whole genome sequencing projects and bioinformatics is used for the annotation of biological data in meaningful ways [3]. Similarly a large collection of problems has been solved by combining the knowledge and abilities of artificial intelligence with bioinformatics approaches for the prediction of genes, studies involving the study of protein interactions, computational systems for drug design, repurposing drugs for better efficiency, next generation sequencing and development of other software. Therefore, both artificial intelligence and machine learning have useful applications in the field of bioinformatics. The proficiency of artificial intelligence can be changed by varying input data. Artificial learning is further classified into generalized and applied branches. Both of them are totally different where applied is involved in the use of machines and algorithms while applied to stimulate the data into expressions similar to the thoughts of humans automatically [4].

A brief history of artificial intelligence was given by Gravey in the form of three systems or paradigms which include; GOFAI having its basis in the 1960s, expert based systems from 1980 to 1990s and machine learning which is prevailed from 2010 to the present. The first technique GOFAI is based on a heuristic technique and was derived from general purpose logical systems. The expert system moved from the general intelligence to human experts which included chemistry and medicines, replicating the descion making processes. This in turn has led to major artificial

intelligence systems in medicines such as MYCIN and more familiar software called TurboTax. These AI algorithms produced similar results but the results were limited and didn't involve the generation of human thinking machines they had promised earlier. Current machine learning has overcome some of the barriers and has predicted a huge amount of data, increased computational power and the revival of neural networks. These algorithms can be trained for exploit further data and thus don't require human labor or programmers for data prediction [5].

Artificial intelligence and computational approaches have revolutionized drug discovery processes in addition to addressing the risks and challenges related to it. Similarly, drug repurposing is a process using which new drugs are developed from existing drugs that provide medical indications beyond their therapeutic use. Drug repurposing offers various advantages over de novo synthesis of drug molecules which involves identification and production of new drug molecules. On the other hand repurposing causes the production of drug in a much earlier time period as compared to new synthesis, provides a much lower rate of failure of drug molecules and leads them to the clinical phase because its ingredients include compounds that have already been tested. In addition drug repurposing allows to the preclinical testing phases in humans and other animal models to be passed away, thereby reduce the overall cost of drug development [6]. If safety testing has been performed then it will display compatibility of dosage with the new indications. Drug repurposing has been discovered by chance on the basis of random testing and exploration. For example, in the market, sildenafil citrate was discovered as a hypertensive drug and then repurposed with Pfizer resulting in the formation of a new drug molecule Viagra for the treatment of erectile dysfunction proved by clinical studies directly from its formulation which in turn provided massive sales in addition to additional health benefits. From the last few years, several computational approaches have been proposed for the repurposing of drug molecules [7]. Popular information for this purpose has been acquired from in silico drug repurposing such as electronic records, gene based information, genetic expression response-based profiles, mapping of the targeted complex molecules and phenotype- based profile assessment. Drug repurposing hubs and repurpose hub were recently surveyed. There is also some literature on drug design, discovery and optimization of lead during development, leading to the development of a totally novel molecule [8].

However, the aim of this review is to use artificial intelligence and machine learning programs that use publicly available data and information. The main point of emphasis is the comprehensive

target based activity of the drugs for the discovery of drug molecules and repurposing where existing drug molecules have some target effects over recently identified target effects for the purpose of new indications [9]. Hence it provides evidence for the further development of drug molecules and their commercialization. This is particularly true for the drugs that are not specific to any target but have the activity for a number of targets showing broad spectrum activity. For example, in some cases cancer off target candidates are available which in addition to having anticancer activity also have potential for the production of new drugs. Here the point to consider is that repurposing us not only to cancer related drugs but also advanced to SARS-CoV novel virus which can be readily applied to patients with COVID-19 [10].

Repurposing was initiated after the phenotypical observation of adventitious drug molecules by polypharmacological activities. For example, we observed the most surprising effects of axinitib which is an inhibitor of the endothelial growth receptor, for treating renal cell carcinoma in both chronic and acute cases. We thought that cancer was caused by the presence of the BCR-ABL1 fusion protein and the drug binds to it. However, the results were surprising based on the computational analysis which showed that this drug binds to the muted form T315I-linked BCR-ABL1 more strongly than the wild form [11]. There are further reports that suggest that the drug may lose its potency when it binds to other molecules and the dosage of the drug is not found to be effective against ponatinib resistant cells. This also showed that these algorithms can also be used to point out the drawbacks before the repurposed drug enters the new phase. Hence artificial learning is now used for the prediction of personalized drug molecules by analyzing the genomic data of patients using certain computational techniques in addition to its diagnosis [11]. These algorithms based on training by AI has more efficiency as compared to the software having no AI training. These include logistic regression, k- nearest neighborhood analysis and linear regression in addition to neural networks in the field of bioinformatics for the computational efficiency in field of Pharmacology, immunology and other biological areas etc. as in Fig.01. The improvement of AI and machine learning in the immunoinformatic and pharmacy has led to the development of clinical immunoinformatic or clinical bioinformatics providing the thorough screening which in turn provide more power to computational simulation in bioinformatics. Now the computational methods have the ability to screen about greater than 100,000 influenza virus using different methods for reaching at a logical conclusion [12].

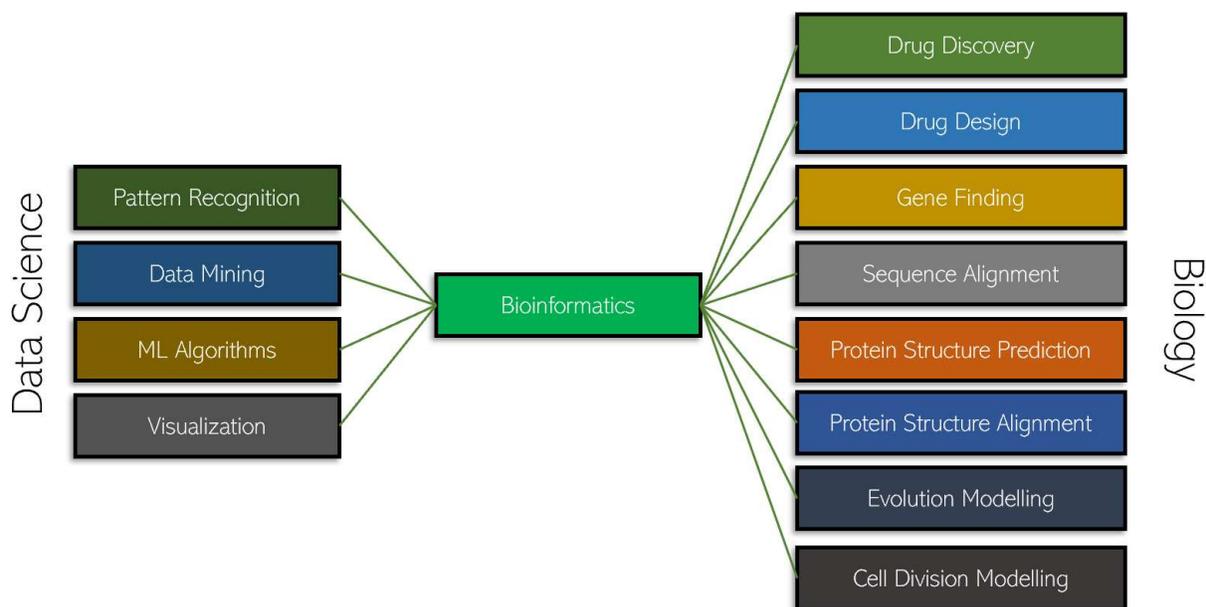


Fig.01. The relationship between data sciences, bioinformatics and biomedical science.

2) Complications in drug discovery:

From the early decades, there has been a long and complex history of the drug development and repurposing. The discovery of the new drug is still a much time taking process requiring about 10-12 years being required for bringing an effective medicine or potent drug molecule from laboratory to market place. Moreover it is very expensive procedure and require about 4-5 billion for discovering a drug molecule which will put financial pressure on the patients which cannot acquire these potentially expensive treatments [13]. As for example following 2020, the cost of drug and treatment in case of cancer is found to increase from. 130 billion dollars to 157 billion dollars. Additionally various drugs have to meet certain properties for approval which include pharmacokinetics, pharmacodynamics, efficacy and toxicity in both the in vivo and vitro studies [14]. The pre-clinical analysis of the drugs continues to check out the safety of a molecule in four phases. Basically, the discovery of drug gets hindered by the efficacy or toxicity profiles during its development. According to the research it has been found that most of the drugs pass the phase 1 of the safety or toxicity trials but are failed in phase 2. As for example in the failure of the drugs in clinical or pre-clinical trials the oncology of cancer related drugs possesses highest failure rate. Recent development in drug discovery allows the development of only one of the drug which targets the disease specifically [15]. However still the success rate is much lower showing that 1

out of every 510 drugs is approved by FDA. Moreover, it was found that only 5% drugs are able to pass the phase 1 of clinical trial. This number decreases further when a drug enters phase 2 and hence suggests much failure rate. For the discovery of drug molecule, it is important for scientists to understand underlying mechanism of disease as for example in case of cancer progression [16]. More than 500 signaling molecules are involved but the drug being discovered only targets one of them. So, there is also need of research in drug targets instead of only focusing on the single target. Hence most of the drugs are being classified on the basis of their pre-clinical studies and these findings are not applicable at clinical levels. Moreover, the lack of the quality in both pharmacokinetics and pharmacodynamics of the drug molecule results in failure of molecules. So the further testing focuses and aim to study the potential of drug molecules from pre-clinical or clinical findings [17].

3) Computational methods for the classifications of variants:

In the recent years, next generation sequencing such as whole genome sequencing has been applied for understanding the genetic basis of the diseases. Using whole exome sequencing WES, the genetic variants in humans can be identified easily which showed that missense variants are responsible for genetic disease in humans. Not all the variants are associated with genetic variants only deleterious one are involved in Mendelian diseases, cancers etc. [18]. Identification of these deleterious variants at a time is a much laborious and time consuming procedure due to its complexity. Hence computational methods have been purposed for solving such problems efficiently by using techniques or methods such as sequence evolutionary techniques, sequence homology and structural similarity data [19]. Most commonly there are three methodological approaches for prediction of variants including;

- **Methods based on sequence conservation:** It usually relies on the conservation of nucleotide base sequence in relation to the multiple sequence alignment at a particular position.
- **Methods based on protein structure prediction:** This is based on the identification of missense variation which in turn create structural differences in the prediction of structures of protein.
- **Ensemble methods:** which take into account both the sequence and structural based properties for calculation or estimation of deleterious variants.

In the case of missense variant identification, almost all of the methods are applied at once being utilized in our studies [20]. Vcards is the computational tool for the identification of genetic variants and has integrated the functional consequences of allele frequencies, computational methodologies and genetic information coded for variants. But it is still difficult to find out the variance in these computational methods because they differ continuously with time [21]. However, a number of studies have compared the missense variation in case of computational methodologies. But they don't employ the use of experimentally obtained datasets. And these studies are based on ROC receiver operating characteristic curves. But the other properties must also be taken into account which included accuracy, efficiency, infectivity, absorption and mechanism in AUC which are not yet been considered [22]. However there also the studies which show that it is used for the prediction of harmful genes in the missense variations. But it us shown that it has to distinguish the pathogenic variants with a high sensitivity using VEST3, REVEL and M-CAP. However, a recent method has employed the use of about 23 computational techniques in this field being listed in table. The prediction of pathogenicity based on these 23 tools can be analyzed from dbNSFP databases. Hence it can be used for the identification and separation of deteriorous genes in the different diseases' conditions especially in the terms of cancer. In addition to it, the information score, genetic score and prediction value are also considered in addition to the results from Varcards databases. The cutoff value was obtained from ANNOVAR, different databases and original studies [23, 24].

4) Artificial Intelligence or machine learning in precision drug discovery:

National Institute of health NIH has highlighted the fact that precision medicine is an emerging strategy for the 'purpose of drug prevention or treatment which also considers the other variations in genetics, libertines and environment. This allows the doctors and physicians to treat and recover diseases more accurately than that of another method employed so far [25]. In order to make this more powerful it requires super mount and fastidious techniques which can be later used in an unfrequented way for trained set of data. The field of artificial learning uses the cognitive ability of the physicians and doctors using biomedical and bioinformatics data to produce fruitful results. Artificial intelligence can be broadly classified into almost three categories which include artificial general intelligence, artificial narrow intelligence and super intelligence [26]. ANI or artificial natural intelligence is still in the process of development and aims to hit the market or research in

next decade. It has the ability to develop new data set, analyze them, to find the correlation among them and draw the meaningful or useful conclusions [27].

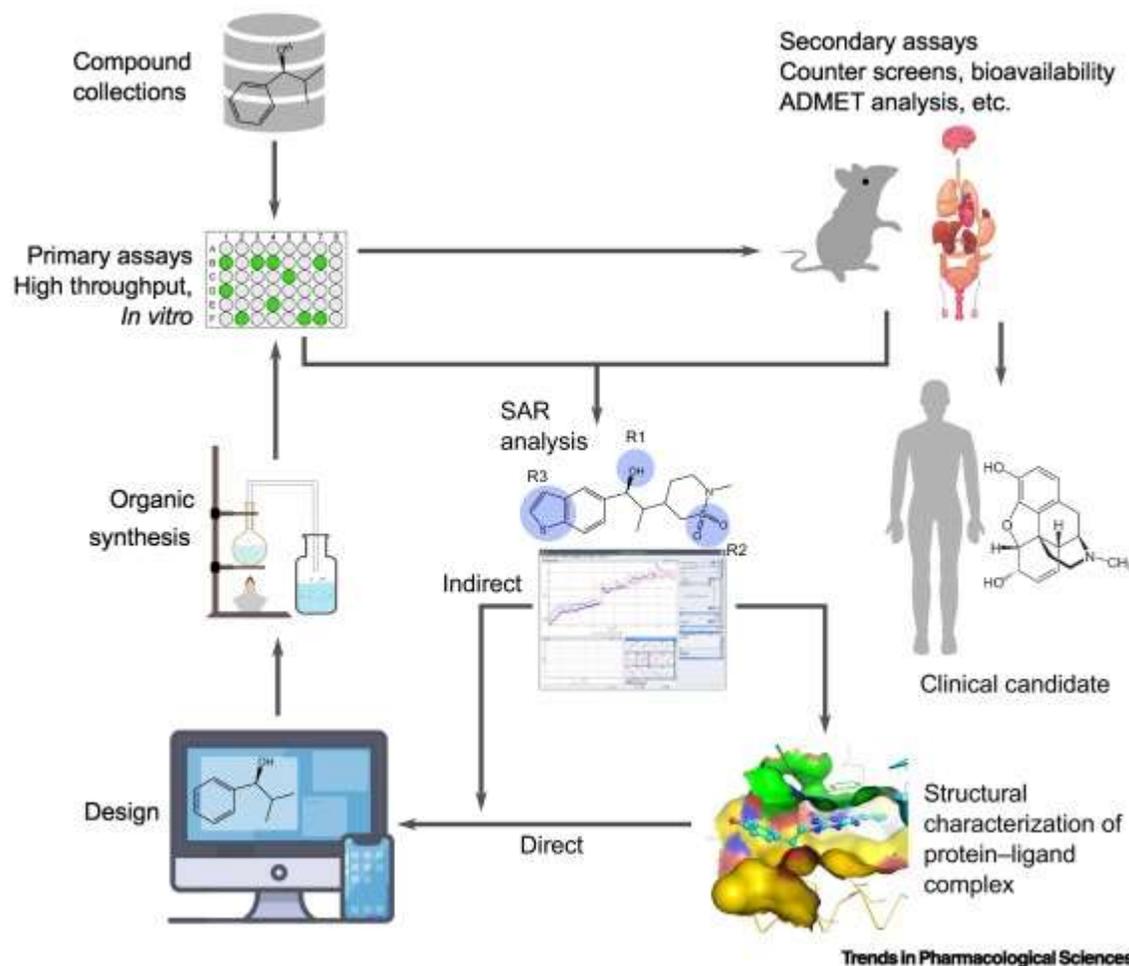


Fig.02. General steps for drug discovery using AI and ML

The most popular pharmaceutical companies are using this deep learning, artificial intelligence and ANI during their development process which aims to identify the unique genetic mutations in a large set of data and thus physicians can use it in various fields of medical science [28]. As we have already discussed that using this technique, we can make drug discovery process more fastidious. Atom wise is used by Biopharma for the identification and prediction of the small structures in 3D basis now a days. The atomwise has used this technique for the identification of active molecules in the case of Ebola virus [29]. Using Artificial Intelligence Company has developed two different drugs which are more effective against Ebola virus as compared to medications which are existed before. Although the method proves to be promising but the

companies need to provide the safety and other consideration overview using a peer review or other methods. The fig 01. Presents the general cycle used by artificial learning in discovery of drug molecule which also uses bioinformatics based data [30].

4.1. Role of AI in drug identification:

Virtual screening has been developed for reducing the cost of high throughput screening and has been used for the better optimization of small molecules acting as the potent drug candidates. The generalization, machine learning and other methods implanting AI has been used for the virtual screening if drug molecule at different points. Virtual screening is further classified into ligand and structure based virtual screening [31]. In the ligand based the binding between legend and structure molecule is analyzed and then predicted based on situation. Based on the knowledge of virtual screening, the structure based analysis can be done using AI methodologies [32]. The correlation existing between the protein –ligand binding free energy and the resulting vector is observed using data driven manner based on the existing experimental data available which also predicts non-linear relationships for getting strong scoring functions. The RF based and SVM based score, ANN and NNS score are all based on AI based non predetermined score which have develop to identify most potential ligands having high accuracy score. The recent advancements in field of AI based non-predetermined scoring had outperformed the existing traditional methodologies for prediction of binding affinity discussed in several researches [33].

For the improvement of AI score five major algorithms were adopted namely SVM, Baysein, deep neural networks and forward ANN reports. Ballsteret al. predicted the techniques for the improvement of these machine learning techniques being employed in increasing or enhancing pre-dominated AI scoring which in turn provides better type of binding between protein and ligand binding complexes. Based on the information, he developed RF based software for predicting protein ligand docking score [34]. Some other scoring functions such as B2B scoring function, SFC scoring, RF-Ichem have also been employed for calculation of docking scores. Regarding comparison of their RF score, the RF score predictions are found to be outstanding which is connected to ISTAR platform used for the large protein –ligand binding dockings [35].

Furthermore, SVM based pipelines have been developed based on both the weaknesses and strength of ligand –protein binding screening. As for example, 18 million compounds are predicted with the activity of c-Met tyrosine kinase inhibitors. PROFILER is the automated workflow

designed by scientists for the identification of the perfect targets which possess highest activity of binding with the targeted molecule [36]. It integrates with two-structural based approaches which include protein ligand binding pharmacophores search and docking and another four structural based approaches which include vector regression affinity based prediction, SVM binary classifications, three dimensional similarity search function and neighboring affinity prediction [37]. In the virtual screen of structure-based prediction, RF score has been found and applied subsequently for the prediction of accurate targets. RF-score VS is actually an enhanced level scoring which are trained on the directions of full decoy datasets. The integration of AI in structure based virtual screening is found to be a promising approach and has been used for improvement of the post processing after the structural based virtual screening by reconsideration of the virtual score calculated with docking algorithms using machine learning models without any consensus scoring. As for examples Auto dock can also be integrated with RF score virtual screening in order to get better performance in VS. Integration of advanced learning and machine-based methods is helpful in the prediction of potential ligands. Hence the involvement of AI and ML in drug discovery has resulted in bringing down the value of false negative and false positive predictions in term of drug discovery process and in future it will also consider the physicochemical properties and structural information of predicted protein [38].

5) Role of machine learning in repurposing of drug by accurate DTI prediction:

Accurate drug target identification is not only overcome the experimentally mapped DTI based methodology but also have the ability to repurpose the drug molecules by targeting the already available drug molecules which are already approved [39]. There are various in silico-based methods available for repurposing of drug molecule using DTI mappings. For instance, Mei et al proposed a multiple learning method for the repurposing of drug based on already known drug targets. During this process each of the drug molecule is considered as class label and the target protein acts as a learning data for training of I2 logistic based regression model [40]. About 89% correct validation is done by stratified multi-labelled cross validation for one of the drug molecule and proposed framework showed 84.5% accurate DTI predictions as compared to others. The fact shows that proposed framework generally is compatible for large frame drugs and require only the information of drug chemical structure and target identification [41]. The recently discovered iDrug method integrates DTI and drug repositioning into a new coherent model termed as cross

network embedding. It provides the principle way of transferring knowledge across various drugs target relationships and using this it predicts accuracy for both DTI and drug disease relationships. The working of iDrug was tested on various datasets containing data from multiple disease types and hence has become applicable for drug repurposing due to several reasons [42].

Molecular transformer drug target interaction MT-DTI is used for getting more accurate results which is pre trained learning based drug targeting model allowed for a number of already present drugs and has been used for the identification of viral proteins effective against SARC-CoV 19. Using this method, atazanavir an antiviral or antiretroviral for HIV treatment and has proved to be the most potent drug having kd value of 94.66 which is comparable to other drugs available for treating SARC-CoV [43].

6) Algorithms for molecular docking and simulation:

Molecular docking is an in-silico approach for the designing of drugs on the basis of their structures as it has the ability to predict the conformational properties of small molecules or ligands attached to drug or appropriate target binding site. But the drawback lies in the fact that it is unable to resolve the molecular 3D structure of many of the drug candidates required for molecular docking or simulation [44]. Moreover, another thing the accuracy of the methods gets decreasing when there are small or insufficient number of ligands available for a drug molecule. Regardless of the limitations, there are number of applications which describe the involvement of AI based tools in molecular docking. For example, thioridazine is an anti-inflammatory compound and has been included in the 1500 FDA approved drugs and also inhibit ikB activity being critical for NF-kB pathway involved in progressing of cancer. Similarly in another experiment, virtual docking has predicted the inhibitory activity of about 1400 FDA approved drug in against to *Pseudomonas aeruginosa* showing inhibitory activity against the various virulence gene expression [45].

Artificial learning is found to be the most accurate method for prediction of structure of the protein from their amino acid sequences [46]. We recently employed Virtual Kinome Profiler which is an efficient platform for distinguishing of chemical similarity of kinome , thus speeding up the drug repurposing process for accurate identification prediction of inhibitors [47]. SVM sequence vector machine algorithm has enabled the prediction of greater than 30M kinase compounds using which we carried out in silico predictions of 150k compounds in term of drug repurposing and optimization of lead compounds. Experimental testing provided the validation of about 15 targets

out of 50 leading to 1.5% increase in precision and decrease of false negative activity to about 1.3% [48].

There are also various examples where various algorithms have been used for drug repurposing in various target classes. CATNIP model for example uses the similarity of molecules on the basis of their target, structure or pathway information. Similarly, another model used the chemical structure-based information for the prediction of 23 FDA approved drug molecules. From them 8 were having cardio protective activities as compared to others [49].

7) Conclusion:

With the emergence of the field, the scientists are in search of new drugs, therapies or treatments having more efficiencies as compared to existing ones. By understanding the underlying mechanism in the progression of disease and the effect of already discovered drug and unveiling of their genetic codes can lead to the new and precision drug therapies which in turn will improve both the health and life of patients. Classical methods used for the discovery of the drugs are found to be time consuming, having more false negative results. In comparison using computational approaches having AI, one can identify the drug more quickly. A number of researches has shown the impact of the drugs on various targets and many of them are in the process of discovery now. In addition to identification of targeted drug or protein, repurposing of drugs provide new molecular targets which can be further used based on their properties also for the prediction of new drug molecule using AI based tools in field of bioinformatics. In future there are expectations that bioinformatics will play a more significant role in the analyzing of data using AI and ML approaches in order to save time and resources at the same time. It will in turn also accelerate the discoveries in the field of biomedical, biological sciences and robotic surgeries.

8) References:

1. Ezziane, Z., *Applications of artificial intelligence in bioinformatics: A review*. Expert Systems with Applications, 2006. **30**(1): p. 2-10.
2. Narayanan, A., E.C. Keedwell, and B. Olsson, *Artificial intelligence techniques for bioinformatics*. Applied bioinformatics, 2002. **1**: p. 191-222.
3. Nicolas, J., *Artificial intelligence and bioinformatics*, in *A Guided Tour of Artificial Intelligence Research*. 2020, Springer. p. 209-264.
4. Bülow, R.D., et al. *How will artificial intelligence and bioinformatics change our understanding of IgA in the next decade?* in *Seminars in Immunopathology*. 2021. Springer.
5. Mitra, S., *Computational intelligence in bioinformatics*, in *Transactions on Rough Sets III*. 2005, Springer. p. 134-152.
6. Das, S., A. Abraham, and A. Konar, *Swarm intelligence algorithms in bioinformatics*, in *Computational Intelligence in Bioinformatics*. 2008, Springer. p. 113-147.
7. Kumar, R., et al., *Prediction of drug-plasma protein binding using artificial intelligence based algorithms*. Combinatorial chemistry & high throughput screening, 2018. **21**(1): p. 57-64.
8. Mellit, A. and S.A. Kalogirou, *Artificial intelligence techniques for photovoltaic applications: A review*. Progress in energy and combustion science, 2008. **34**(5): p. 574-632.
9. Chen, S.H., A.J. Jakeman, and J.P. Norton, *Artificial intelligence techniques: an introduction to their use for modelling environmental systems*. Mathematics and computers in simulation, 2008. **78**(2-3): p. 379-400.
10. Laughton, M., *Artificial intelligence techniques in power systems*. 1997.
11. Fethi, M.D. and F. Pasiouras, *Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey*. European journal of operational research, 2010. **204**(2): p. 189-198.
12. Mellit, A., et al., *Artificial intelligence techniques for sizing photovoltaic systems: A review*. Renewable and Sustainable Energy Reviews, 2009. **13**(2): p. 406-419.
13. Dilek, S., H. Çakır, and M. Aydın, *Applications of artificial intelligence techniques to combating cyber crimes: A review*. arXiv preprint arXiv:1502.03552, 2015.

14. Elbadawi, M., S. Gaisford, and A.W. Basit, *Advanced machine-learning techniques in drug discovery*. Drug Discovery Today, 2020.
15. Vasant, P., *Handbook of research on artificial intelligence techniques and algorithms*. 2014: IGI Global.
16. McLaren, B.M., O. Scheuer, and J. Mikšátko, *Supporting collaborative learning and e-discussions using artificial intelligence techniques*. International Journal of Artificial Intelligence in Education, 2010. **20**(1): p. 1-46.
17. Nawari, N., R. Liang, and J. Nusairat, *Artificial intelligence techniques for the design and analysis of deep foundations*. Electronic Journal of Geotechnical Engineering, 1999. **4**: p. 1-21.
18. Kohli, A., et al., *Concepts in US Food and Drug Administration regulation of artificial intelligence for medical imaging*. American Journal of Roentgenology, 2019. **213**(4): p. 886-888.
19. Shi, F., et al., *Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19*. IEEE reviews in biomedical engineering, 2020.
20. Rahmanifard, H. and T. Plaksina, *Application of artificial intelligence techniques in the petroleum industry: a review*. Artificial Intelligence Review, 2019. **52**(4): p. 2295-2318.
21. Zhavoronkov, A., *Medicinal Chemists versus Machines Challenge: What Will It Take to Adopt and Advance Artificial Intelligence for Drug Discovery?* Journal of Chemical Information and Modeling, 2020. **60**(6): p. 2657-2659.
22. Sfetsos, A. and A. Coonick, *Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques*. Solar Energy, 2000. **68**(2): p. 169-178.
23. Zhou, Y., et al., *Artificial intelligence in COVID-19 drug repurposing*. The Lancet Digital Health, 2020.
24. Fleming, N., *How artificial intelligence is changing drug discovery*. Nature, 2018. **557**(7706): p. S55-S55.
25. Lamberti, M.J., et al., *A study on the application and use of artificial intelligence to support drug development*. Clinical therapeutics, 2019. **41**(8): p. 1414-1426.
26. Martinez-Romero, M., et al., *Artificial intelligence techniques for colorectal cancer drug metabolism: ontologies and complex networks*. Current drug metabolism, 2010. **11**(4): p. 347-368.

27. Yang, X., et al., *Concepts of artificial intelligence for computer-assisted drug discovery*. Chemical reviews, 2019. **119**(18): p. 10520-10594.
28. Jiménez-Luna, J., F. Grisoni, and G. Schneider, *Drug discovery with explainable artificial intelligence*. Nature Machine Intelligence, 2020. **2**(10): p. 573-584.
29. Díaz, Ó., J.A. Dalton, and J. Giraldo, *Artificial intelligence: a novel approach for drug discovery*. Trends in pharmacological sciences, 2019. **40**(8): p. 550-551.
30. Zhu, H., *Big data and artificial intelligence modeling for drug discovery*. Annual review of pharmacology and toxicology, 2020. **60**: p. 573-589.
31. Mohanty, S., et al., *Application of Artificial Intelligence in COVID-19 drug repurposing*. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2020.
32. Barrett, S. and W. Langdon, *Advances in the application of machine learning techniques in drug discovery, design and development*, in *Applications of Soft Computing*. 2006, Springer. p. 99-110.
33. Zhavoronkov, A., *Artificial intelligence for drug discovery, biomarker development, and generation of novel chemistry*. 2018, ACS Publications.
34. Zhong, F., et al., *Artificial intelligence in drug design*. Science China Life Sciences, 2018. **61**(10): p. 1191-1204.
35. Paul, D., et al., *Artificial intelligence in drug discovery and development*. Drug Discovery Today, 2020.
36. Stephenson, N., et al., *Survey of machine learning techniques in drug discovery*. Current drug metabolism, 2019. **20**(3): p. 185-193.
37. Torres-Carrión, P., et al. *Application of techniques based on Artificial Intelligence for predicting the consumption of drugs and substances. A Systematic Mapping Review*. in *International Conference on Applied Technologies*. 2019. Springer.
38. Chan, H.S., et al., *Advancing drug discovery via artificial intelligence*. Trends in pharmacological sciences, 2019. **40**(8): p. 592-604.
39. Zhavoronkov, A., Q. Vanhaelen, and T.I. Oprea, *Will Artificial Intelligence for Drug Discovery Impact Clinical Pharmacology?* Clinical Pharmacology & Therapeutics, 2020. **107**(4): p. 780-785.

40. Weinstein, J.N., et al., *Predictive statistics and artificial intelligence in the US National Cancer Institute's drug discovery program for cancer and AIDS*. *Stem cells*, 1994. **12**(1): p. 13-22.
41. Kumar, R., et al., *Prediction of human intestinal absorption of compounds using artificial intelligence techniques*. *Current drug discovery technologies*, 2017. **14**(4): p. 244-254.
42. Koromina, M., M.-T. Pandi, and G.P. Patrinos, *Rethinking drug repositioning and development with artificial intelligence, machine learning, and omics*. *Omics: a journal of integrative biology*, 2019. **23**(11): p. 539-548.
43. Bajorath, J.r., et al., *Artificial intelligence in drug discovery: Into the great wide open*. 2020, ACS Publications.
44. Levin, J.M., et al., *Artificial intelligence, drug repurposing and peer review*. *Nature Biotechnology*, 2020. **38**(10): p. 1127-1131.
45. Álvarez-Machancoses, Ó. and J.L. Fernández-Martínez, *Using artificial intelligence methods to speed up drug discovery*. *Expert opinion on drug discovery*, 2019. **14**(8): p. 769-777.
46. Wang, L., et al., *Artificial intelligence facilitates drug design in the big data era*. *Chemometrics and Intelligent Laboratory Systems*, 2019. **194**: p. 103850.
47. Green, C.P., O. Engkvist, and G. Pairaudeau, *The convergence of artificial intelligence and chemistry for improved drug discovery*. 2018, Future Science.
48. Bender, A. and I. Cortes-Ciriano, *Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data used for AI in drug discovery*. *Drug Discovery Today*, 2021.
49. Gertrudes, J., et al., *Machine learning techniques and drug design*. *Current medicinal chemistry*, 2012. **19**(25): p. 4289-4297.