Life Science Informatics Publications
Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences
Journal Home page http://www.rjlbpcs.com/
RJLBPCS
ISSN 2454-6348

Original Review Article **DOI: 10.26479/2023.0904.03**

DESIGN OF POTENTIAL DRUGS AND VACCINES FOR COVID- 19 BY THE APPLICATION OF MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Indrani Sarkar[1]*, Sudeshna Sarkar[2]

1. Department of Basic Science and Humanities (Physics), Narula Institute of Technology, 81, Nilgunj Road, Agarpara, Kolkata 700109, West Bengal, India.
2. Department of Tropical Medicine, Calcutta School of Tropical Medicine, 108, Chittaranjan Avenue, Calcutta Medical College, College Square, Kolkata 700073, West Bengal, India.

ABSTRACT: Computational methods like neural network and genetic algorithm can be used to speed up the drug discovery process. Machine learning method like artificial neural networks is widely used in pharmaceutical industry. Design of combinatorial libraries, docking of drug molecules to targets like proteins and DNA are a few applications. Neural networks can be applied for applications like Association, Classification (clustering), Transformation (different representation) and modeling. Quantitative Structure-Activity /Property Relationship (QSAR/QSPR). QSAR/QSPR correlates topological, electronic and quantum properties of compounds with their biological activities. ANNs have been found to be effective in building QSAR/QSPR models to predict the activities of new compounds. Virtual Screening (VS) method "screen" compounds with known chemical motifs called "pharmacophores" amid millions of other compounds in a data base. These screened compounds are potential drugs and can be tested experimentally for their biological activities. Some more applications of ANN are: data analysis, comparison/classification of drug libraries, and study of HIV-1 reverse transcriptase, gene prediction and homology searches in protein. Genetic algorithms (GAs) are stochastic optimization methods. A QSAR model can be made by variable selection, PLS (partial least squares) and cross validation using GA. Pharmacophore modeling is done by comparing some important electronic and 3D structural features required for a potent group of ligands/drugs when the receptor or target is unknown. This article discusses immense application of ANNs and GAs in the drug discovery process. The future prospect of Artificial Intelligence in drug and vaccine design, COVID-19 management and prediction is also discussed.

Keywords: Computer Aided Drug Design, Covid-19, Machine Learning, Artificial Intelligence, Molecular Modeling, Covid-19 Vaccine Design.

Article History: Received: July 20, 2023; Revised: July 28, 2023; Accepted: August 08, 2023.

Corresponding Author: Dr. Indrani Sarkar*, Ph.D.
Department of Basic Science and Humanities (Physics), Narula Institute of Technology, 81, Nilgunj Road, Agarpara, Kolkata 700109, West Bengal, India Email Address: indrani.sarkar@nit.ac.in

1. INTRODUCTION

Classical method for drug discovery is time consuming, expensive and has very low success rate. It involves the following processes: at first suitable drug targets are identified, then the drug target is validated, suitable molecules are extracted from data bank, lead molecules are optimized, and sent for preclinical and clinical studies. The failure rate by classical method is still very high at clinical study level (10-13%). The cause of drug failure at clinical stage is mostly due to ADME/Tox properties, ADMET stands for absorption, distribution, metabolism, excretion, and toxicity which are essential for a drug to be approved. The advent of the recent computer-aided drug discovery (CADD) technique has helped to minimize the costs and failures of drug discovery. CADD (also known as rational drug design) provides detail information about the binding affinity between target macromolecule and ligand at molecular level. In current times artificial intelligence (AI) and machine learning (ML) methods are being used in screening big data. Two important methods involved in CADD are structure-based drug design (SBDD) and ligand-based drug design (LBDD). The disease that shook the whole world in 2019 is caused by a virus which is a sub class of large coronavirus family. Its nucleotide sequence has high similarity with other coronavirus species like Severe Acute Respiratory Syndrome Corona Virus (SARS-CoV) and Middle East Respiratory Syndrome Corona Virus (MERS-CoV)[23]. SARS-CoV-2 has a single-stranded RNA with 30,000 base pairs. It utilizes human cellular mechanism for infection processes. It enters the human cell, replicates and synthesize important proteins. Table 1 shows the potential key target proteins responsible for the infection and replication process inside the human body. Drugs targeting these proteins are potential lead compound to treat COVID-19 infection. SARS-CoV-2 structure has four major structural proteins. Out of these three proteins are used as favorite targets for vaccine design: spike protein(S), membrane protein (M) and envelope protein (E). These three proteins are found on the envelope. The nucleocapsid (N) protein lies within the ribonucleoprotein core (Fig.1). [34]. S proteins recognize the host receptor and starts virus entry. M proteins give shape to the virus envelope. E proteins are essential for CoV infection. N proteins bind along the viral RNA genome. Additional non-structural proteins (nsp 1–16) and accessory proteins have separate roles (Table1, Table2), The nonstructural proteins produce sub genomic RNAs [13, 8]. The nonstructural and

structural proteins, together can be utilized as targets for the design of antiviral drugs/vaccines [13,23].

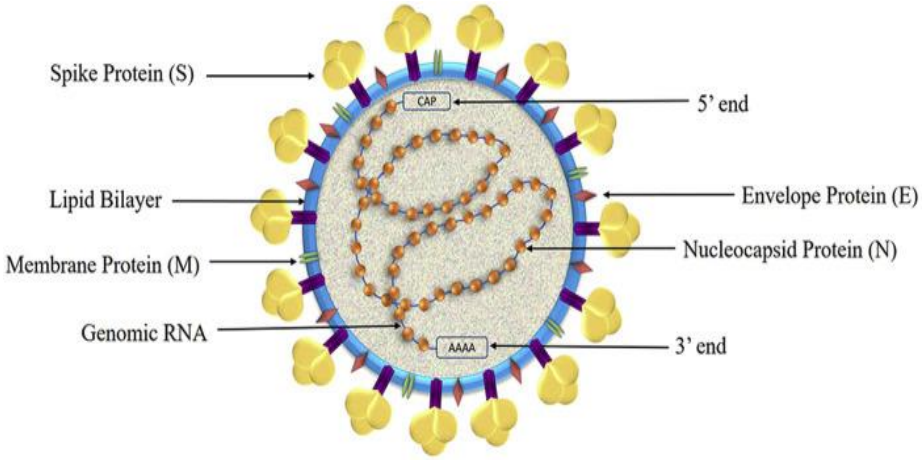

Figure 1. Major structural proteins of Covid 19 virus

Table1: Viral Proteins

Name of proteins	Role of the protein in viral life cycle	Drug used against it
Main protease3CLpro (3CLpro)	It proteolyzes the polyprotein	Lopinavir, Ritonavir
papain-like protease PLpro (PLpro)	It proteolyzes the polyprotein	Ritonavir Lopinavir
RNA-dependent RNA polymerase (RdRp)	Helps to replicate viral gene	Remdisivir, Ribavirin
Spike glycoprotein (S protein)	It helps the virus to bind with the host cell receptor ACE2	Arbidol
Tran membrane protease, serine2 (TMPRSS2)	It helps the virus to bind with the host cell receptor ACE2	Camostat
Angiotensin- converting enzyme2 (ACE2)	It helps the virus to bind with the host cell receptor ACE2	Arbidol

Table 2: Drugsused againstSars-CoV-2

Drugs acting on virus	Remdesivir, Favipiravir, Darunavir, Ribavirin, Lopinavir, Ritonavir, Arbidol, Azithromycin, Nitazoxamide, Elbasvir, Sofosbuvir, Bictegravir, Ivermectin, Prulifloxacin, Cepharanthine, Nafamostat, Nelfinavir, Doxycycline
Drugs acting on human	Chloroquine, Hydroxychloroquine, ARBS, Statins, Interferon β, Tocilizumab, Dexamethasone, Ruxolitinib, Baricitinib

Different vaccine strategies

Seven strategies of vaccine design are shown in Fig 2. The technologies involved are as follows

Inactivated virus vaccines: Hereinactivated virus particles are used as immunogens.

Virus-like nanoparticle vaccines do not possess the viral genome.

Protein subunit vaccines: Here key viral proteins are manufactured in vitro by using bacteria or another cell. Most of the SARS vaccines follow this strategy

Virus-vectored vaccines: Here a harmless adenovirus is used to mimic the infection of the SARS virus

mRNA and DNA vaccines: DNA and mRNA vaccines use part of the genetic code of the virus's own gene to stimulate an immune response. The first time mRNA vaccine for COVID-19 was co-developed by Pfizer and BioNTech.

Live-attenuated vaccines

Live attenuated vaccines are a weaker version of living active viruses. It does not cause serious disease in healthy people. One example of a live attenuated vaccine is the measles, mumps, and rubella vaccine (MMR)

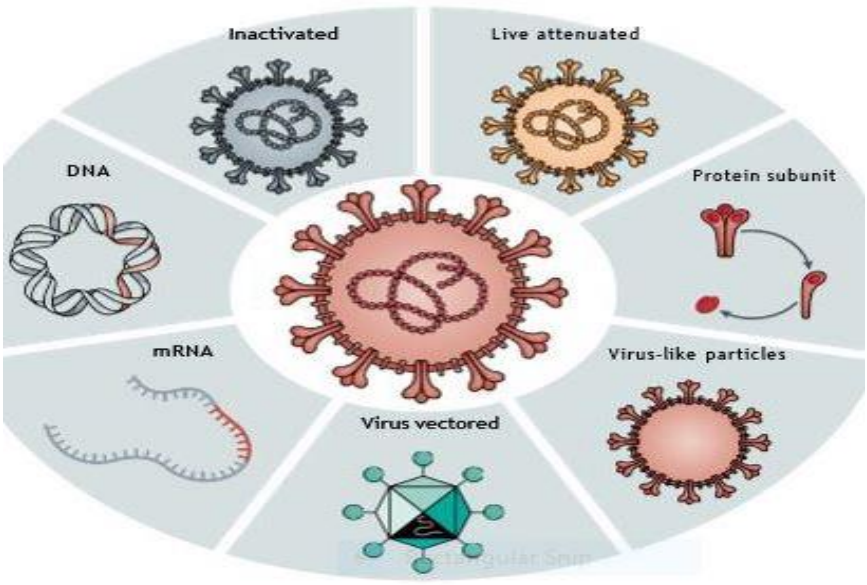

Figure 2. Vaccine design technologies against Covid 19 virus (Dai & Gao, February 2021)

2. MATERIALS AND METHODS

2.1 Data Mining used in Biological Information

Data mining is one of the most exciting areas in modern biology. There is a vast pool of biological databases. Machine learning is a family of programs that adopt their behavior with experience. Programs are made to learn or to be trained; Machine learning methods are of two types: supervised and unsupervised learning

Supervised and unsupervised learning

When a learning algorithm is given a set of labeled examples that is used to train the program (training set) and then then the program is tested or validated on a set of unlabelled examples (test set), it is called supervised learning. When data is available, but the correct labels for each example are not known, it is called unsupervised learning. The learning algorithm is applied on the data and finds some pattern that helps to understand the data. Clustering analysis is a kind of unsupervised learning. It clusters similar data in the same class.

Decision trees: A decision tree has a list of questions with yes or no answers. This is arranged hierarchically to produce a decision. In biology, tree classifications are used to find gene splice sites.

Neural networks: These algorithms also do pattern recognition and classification. It was developed in the 1940s as a mathematical model mimicking human memory. Neural networks are represented as nodes (variables) connected by weighted functions. The nodes have their input and output connections to other nodes. A neural network receives an input (for example, an amino acid or gene sequence) at the first layer which sets the values of the nodes (the input layer) (Fig 3). These values are transferred to the next layers (by transfer functions) up to the output layer. This output

layer pattern gives the output of the network. Neural networks are frequently used in bioinformatics to find secondary structures of proteins from their sequence or to build a homology model of a protein by using its sequence and a template structure from the protein data bank (PDB). Examples of such neural networks are PHD and PSIPRED. These are used for secondary structure prediction, visualizing the macromolecular structure, and computing structural parameters like bond lengths, bond angles, torsion angles, etc. The GRAIL gene finder is used for database searching, and sequence alignment of amino acids in case of homologous proteins.

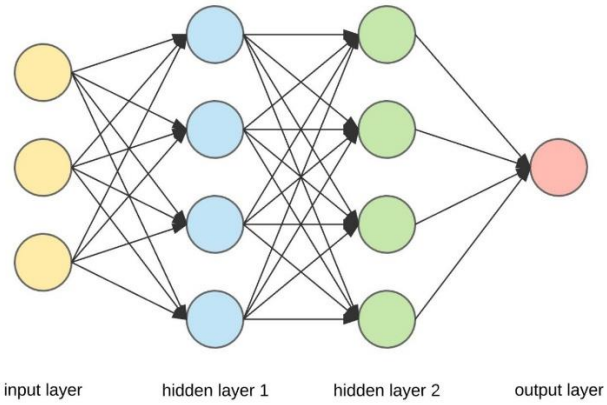

Figure 3. Workflow of an Artificial Neural Network

2.2 Genetic algorithm

Genetic algorithms search a large number of solutions and select the optimized or best one. The best solution is determined by a cost function or fitness function. These models are based on the biological ideas of genetics. In a genetic algorithm, a number of solutions generated at random (Fig. 4) are known as chromosomes. These are exchanged and recombined. The recombined results are evaluated using the fitness function. The highest-scoring chromosomes are propagated to the next generation. This loop continues until a suitable solution is found. Genetic algorithms are used in molecular simulation, molecular docking, protein folding, and QSAR studies.

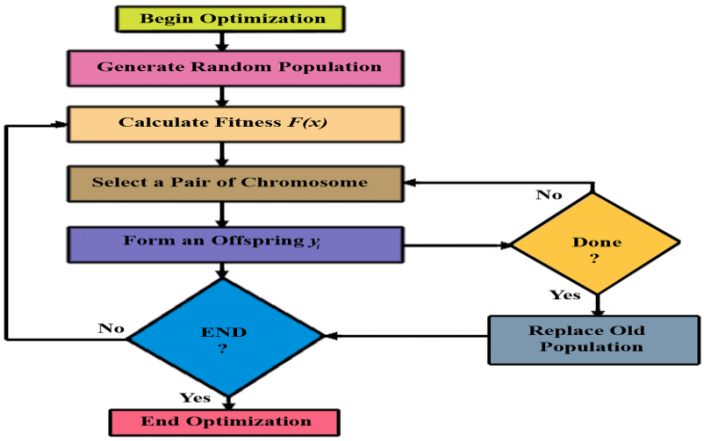

Fig 4: Workflow of Genetic Algorithm

Support vector machines: SVMs are supervised classifiers. They have been used in many standard computational biology problems like structure and function prediction, gene finding, protein sequence classification, etc. The pharmacokinetic andpharmacodynamic data of broad spectrum of antiviral drugs approved by FDA were used as trials to treat Covid 19 [26]. The completegenomesequence of SARS-CoV-2 has been determined and available in Data Bank. Three dimensional structures of some important viral proteins have been determined by x-ray crystallography, nuclear magnetic resonance (NMR) and molecular modeling studies. The inhibition mechanism of the proteins as revealed by molecular modeling and dynamics simulation has helped to design suitable inhibitor drugs. Some in silico methods are discussed here.

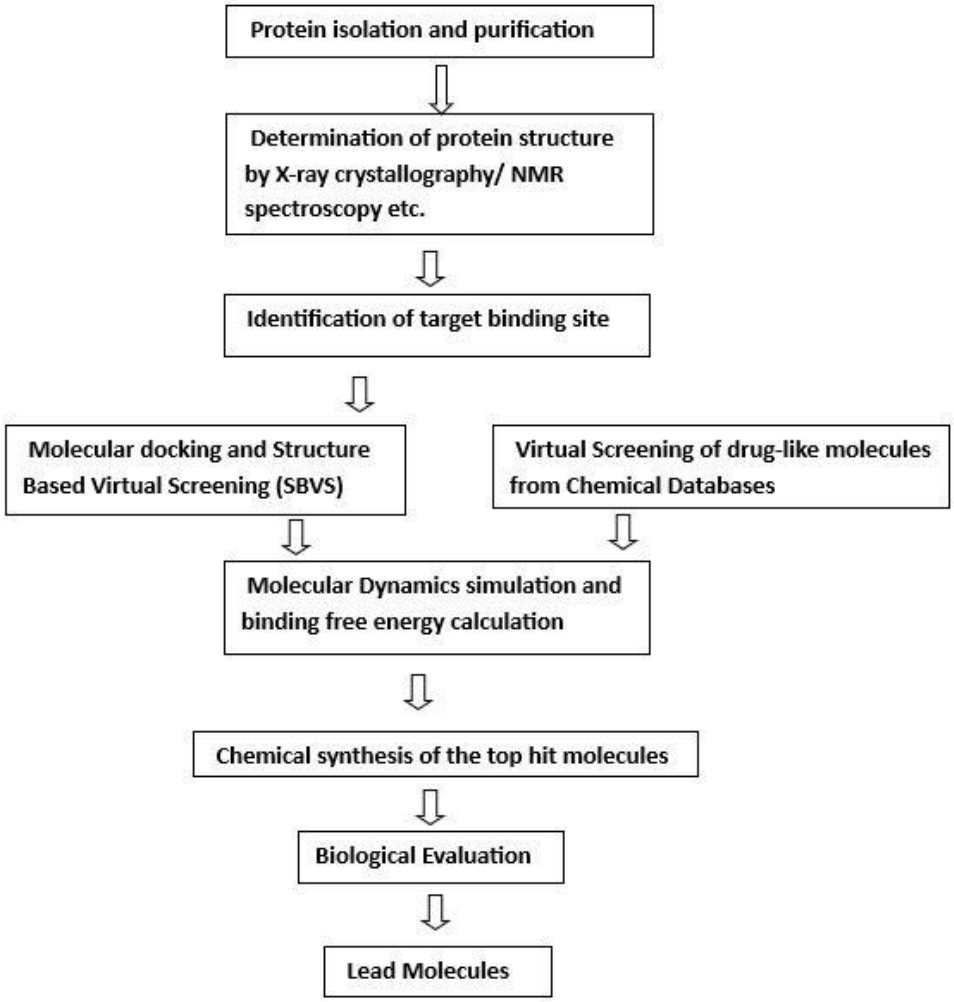

Fig 5: Steps in Computer-Aided Drug Discovery

2.3 Structure-Based Drug Design (SBDD)

This method is applied when the protein crystal structure is available and the important amino acids within the active site cavity are identified [5]. Different steps in SBDD are shown in Fig 5. Molecular docking followed by energy minimization and molecular dynamics (MD) simulations [20]. are carried out as a part of molecular modeling. The molecular dynamics trajectory gives information about binding free energy of protein-inhibitor complex, molecular interactions between target and ligand andchanges occuring within the active site upon binding with a ligand [16,44]. Several drugs found inthemarket have been discovered by using SBDD shown in Table 3.

Table 3. Drugs discovered by Structure-Based Drug Design (SBDD) and Structure-Based Virtual Screening (SBVS)

Drug	**Target/disease**	**The method used/ Company**
Thymidylate synthase inhibitor Raltitrexed	inhibit the enzyme thymidylate synthase and have potential as anticancer chemotherapy.	SBDD,Agouron (La Jolla, CA, USA)
TopoisomeraseIIandIVinhibitor	important drugs used in the therapy of many neoplasms including breast cancer, lung cancer, testicular cancer, lymphomas and sarcomas.	SBVS
Norfloxacin	This antibioticis used to treat urinary tract infection	SBVS
Dorzolamide	This compound inhibits carbonicanhydrase. It is usedto treatglaucoma	pharmacophore modeling, Merck Sharp and Dohme (Harlow, UK)
antituberculosisdrug,isoniazid	It inhibits bacterial protein enoyl-acyl-ACP reductase (InhA)	SBVS, pharmacophoremodeling
Flurbiprofen	It inhibits cyclooxygenase-2. It is used to treat rheumatoidarthritis, osteoarthritis etc	molecular docking approachetc.
Imatinib	BCR–ABL	Novartis (Basel)

Ro46624	Thrombin	Roche (Basel, Switzerland)
Zanamivir	Neuraminidase	Biota (Melbourne, Australia)
Saquinavir	HIV Protease	Roche (Welwyn, UK)
Ro466240	Thrombin	Roche (Basel, Switzerland)

2.4 Some important steps of SBDD are discussed below

Modeling of the 3-dimensional Protein Structure

The RCSB Protein Data Bank (PDB) contains crystal and NMR structures of a large number of proteins and nucleic acids. Due to the complexity of X-ray crystallography and NMR techniques structures of many macromolecules cannot be determined. A number of computer modeling techniques are being used to model the three-dimensional structures of unknown proteins. Examples are homology modeling, threading, and abinitio modeling. In homology modeling a template protein structure is taken to build the 3D model of the unknown protein. Here the sequences of the unknown protein and the template proteins are aligned to find homology between them. The ab initio-modeling is used when the target protein to be modeled does not have any template or similar structures in the data bank.

Identification of the Binding/Active Site of the target molecule

The active site is the region in a protein where the substrates/inhibitors bind. So, knowledge about the positions of the amino acids in the active site is required for docking ligands. There are many freely available software and web server using different machine-learning methods (Table 4) for molecular docking.

Docking the ligand in the active site

Molecular docking method issued to dock a ligand into the active site of the target macromolecule [15]. The computation is done by Monte Carlo (MC) and molecular dynamic (MD) protocol. An energy scoring function is required for all molecular docking programs. The scoring function represents the affinity of binding between the ligand and macromolecule. Force field calculates the actual interaction forces (hydrophobic, hydrophilic, van der Waal's) and finds an energy optimized structure. Some force fields are CHARMM, Amber, OPLS etc.

Table 4. Molecular docking programs

Tools/ Webservers	Method used
AutoDock	It uses Lamarckian genetic algorithm, simulated annealing search, to dock the ligand in the active site
AutoDock Vina	It uses sophisticated gradient optimization method for predicting binding mode and is more accurate than AutoDock
GOLD (genetic optimization for ligand docking)	It uses genetic algorithm for searching docking conformation
CDOCKER	CHARMM based docking software

Preparation of Library of screened compounds

Chemical compounds are selected from small molecular chemical databases such as ZINC, PubChem, ChEMBL, Drug Bank, and ChemSpider. Millions of compounds from these databases are screened by the virtual screening method. For further screening "Lipinski's rule of five" (which is also known as Pfizer's rule of five), ADMET properties, carcinogenicity, and different pharmaco-kinetic studies are carried out. Drug likeness of compounds is assessed by measuring logP values, and molecular properties like the number of H-bond donors and acceptors. ADMET measures the absorption, distribution, and metabolism of drugs in the body. It also measures the amount of drug excreted. Finally, the measurement of the toxicity of drugs is essential for regulatory approval.

Molecular Dynamic (MD) Simulation

Molecular energy minimization followed by MD simulations is used to examine the dynamics of the interaction of the target with the ligand. The dynamics trajectory also reveals details of binding, unbinding, and changes in the target [9]. Some MD simulation programs with suitable simulation systems are shown in (Table5).

Table 5. Some MD simulation programs

Software	Simulation system
GROMACS	Proteins, lipids, carbohydrates, nucleic acids
AMBER	Proteins, carbohydrates, nucleic acids
CHARMM	Proteins, lipids, carbohydrates, nucleic acids
NAMD	Proteins, lipids, carbohydrates, nucleic acids
Desmond	Proteins, lipids

2.5 Ligand-Based Drug Design

This method is applied in CADD when the crystal structure of the target macromolecule is not available in the data bank. In this approach, a number of compounds that are known to have inhibitory activities against a specific target are taken. The biological activity of these compounds is experimentally measured. Now the common structural features among these compounds are screened by using different methods known as pharmacophore modeling and quantitative structure-activity/property/toxicity (QSAR/QSPR/QSTR) relationships. These techniques use genetic algorithms, neural networks, principal component analysis, and artificial intelligence (AI). These algorithms correlate the structural features of a compound with its biological activities as structural similarities produce similar biological function

Pharmacophore Modeling

A pharmacophore is a structural element in the compound which is responsible for its biological/inhibitory activity. A pharmacophore has some notable chemical features like presence of hydrogen bond donor acceptor atoms, hydrophobic site, aromatic rings, ionizable groups [40]. Pharmacophores can be identified by alignment in the pharmacophore based virtual screening [21]. Some frequently used software which build the pharmacophore model automatically are Catalyst, Pharma Gist, Pharmer, PHASE, ZINCPharmer, Ligand Scout, GALAHAD, and Pharm Mapper. The pharmacophore model is validated by using a known set of ligands called test set. When the receptor three-dimensional structure is not available, pharmacophore method is the best method to screen active ligands.

2.6 Quantitative Structure-Activity Property/Toxicity Relationships (QSAR/QSPR/QSTR)

These studies relate the bioactivity of the compounds with their molecular structure or descriptors. Descriptors are molecular properties like molecular weight, volume, electron positivity, electro-negativity, etc. A small change in molecular structure (i.e., by the introduction of a side chain) can bring significant change in its biological activity. This method is widely used in CADD to identify the lead molecule. To generate a QSAR model one has to collect a sufficient number of datasets obtained by one type of experiment. The compound data set is then divided into two sets, training, and test set. The training set data to train the model. The test set data is used for checking (valediction) the applicability of the model. The statistical model is built by correlating the biological activity with the descriptors. The descriptors are taken as independent variables and the biological activity is taken as the dependent variable. An MLR (Multiple Linear Regression) models is built as $y=a_1x_1+a_2x_2+a_3x_3$....where y is the biological activity, x_1, x_2, x_3 are molecular descriptors, and a_1, a_2, a_3 are coefficients of the MLR model. The requirements of a reliable QSAR model are (i) sufficient biological activity data for a group of compounds using same experimental procedure (like IC_{50} data of a protease) and (ii) the total data set is divided into training and test set compounds (iii) there should be no correlation among the descriptors to avoid over fitting of the data (iv) the efficiency of prediction of the model should be

checked by internal and external validation (v) only those descriptors which can describe the biological activity properly are chosen. Machine learning methods like genetic algorithms and neural networks are applied in this case. [17,19]. Hypo Gen module of Catalyst [18], PHASE [10], Comparative Molecular Feld Analysis (CoMFA) [6], and Comparative Similarity Indices Analysis (CoMSIA) are a few tools for QSAR studies. Molecular descriptors are calculated by using software like PaDEL Descriptor, E DRAGON, CODESSA PRO, and MOPAC. Statistical methods like linear and multiple linear regression, partial least squares, and principal component analysis are used in linear QSAR methods. Artificial neural networks and Bayesian neural nets [24] are used in nonlinear QSAR methods [6].

2.7 Application of Artificial Intelligence in Drug Discovery

Artificial intelligence (machine intelligence) allows the computer to learn from given data. AI has vast applications in bioinformatics to predict the bio-activities and toxicities of compounds. AI has been able to predict very complex problems like protein folding and protein-protein interaction. It is also applied for de novo drug design [41]. Two methods used widely in CADD are machine learning (ML) and deep learning (DL) [25]. ML algorithms use support vector machine (SVM) [7], Random Forest (RF) [3] and Naive Bayesian (NB) [32]. A deep neural network (DNN) is an example of a deep learning method. High-throughput screening (HTS) used in the pharmaceutical industry produces Big Data. AI methods can work with this high-volume data and predict both the efficacy and side effects of drugs in humans without going through rigorous laboratory experiments. Deep learning was first used in bioinformatics in 2012 [22]. Deep learning models produce better results than machine learning methods.

3. DISCUSSION

3.1Application of CADD in Covid 19 Research:

The complete genome sequence of SARS-CoV-2 is available now. The three-dimensional molecular structures of important viral proteins determined by crystallography and NMR spectroscopy are available in the data banks. It is already mentioned that the structural proteins shown in Fig 1 are important drug targets of SARS-CoV2. The entry of SARS-CoV-2 into human cells is assisted by a protein called angiotensin converting enzyme-2 (ACE-2). The S protein is responsible for ACE2-mediated virus attachment and membrane fusion. Some amino acids in the viral spike protein enhance ACE2 binding. The Nucleocapsid protein (N) helps the viral entry and further processing inside human cells. The E protein helps in the formation of viral assembly. The M protein facilitates virion production. Thus, all structural proteins play some role in the viral infection process. So, an insight into their structure and function will help to design therapeutic agents.

The non-structural proteins are also attractive targets for therapeutic agents. So, the availability of complete three-dimensional structures of these proteins will help the CADD process immensely in Covid-19 research.

3.2 Some ongoing CADD research on Covid 19: Case Studies

Till date a large number of research articles have been published on the application of CADD techniques on the viral proteins. In 2020 a deep-learning trained drug-receptor interaction protocol was used for 3C-like protease [4]. The model called molecular transformer predicted a number of FDA approved antiviral drugs. Some of these drugs are remdesivir, atazanavir, ritonavir, dolutegravir and efavirenz. A three-dimensional structure of nsp-14 (a non-structural protein of SARS-CoV-2) was built by homology modeling from its sequence [11,33]. A number of antiviral phytochemicals were predicted based on molecular modeling studies [32]. Elfiky (2020) [12] built homology model of SARS-CoV-2 RNA dependent RNA polymerase (RdRp) protein. Drugs like remdesivir, favipiravir, cefuroxime, tenofovir, hydroxychloroquine, sofosbuvir, ribavirin and galidesivir were docked with RdRp. Subsequent molecular dynamics simulation revealed the interaction mechanism of these drugs with the protein. These drugs were taken for clinical trial [12]. Das et al. (2020) [11] studied the inhibitory activity of some natural compounds by modeling against SARS-CoV-2 main protease. Wahedi et al. (2021) [42] showed that viral main protease and human ACE-2 complex formation can be disrupted by piceatannol and resveratrol. These two compounds are potential anti-COVID-19 drugs. They performed molecular dynamics simulation of the docked complex (PDB entry 6LU7). Skariyachan et al. (2020) [37] checked the inhibitory properties of 15 compounds and six FDA approved drugs (chloroquine, hydroxychloroquine, favipiravir, lopinavir, remdesivir, and ritonavir). It was found that ritonavir and lopinavir are better inhibitors than other drugs [31,36,37,38]. Considering the severity of Covid 19 situation there is a need for vaccines and drugs in less time. The CADD technique has accelerated this process. It has also predicted the use of already existing FDA-approved drugs with known side effects. The high degree of mutation capacity of the SARS-CoV-2 genome can pose difficulty in the design of vaccines and treatment. This in turn will affect the mechanism of drug binding with the receptor. So, CADD can be very helpful in predicting the effects of mutation on the receptor protein. All these can be visualized through molecular modeling and dynamics simulation studies. Still CADD has some limitations because the lead molecules screened from the database by the virtual screening method are not validated by preclinical and clinical studies. So, the prediction accuracy is questioned. Again, due to lack of sufficient data AI also cannot be very helpful.

4. CONCLUSION

4.1 Future prospect: **Use of Big Data, Artificial Intelligence and Machine Learning in fighting Covid 19: Algorithms and Case studies**:

Many pharmaceutical laboratories started developing vaccines using different vaccine technologies immediately after the outbreak of Covid-19 pandemic. Each technology has its own advantages and disadvantages. The use of AI, ML, and big data in vaccine development has shortened the time of discovery. But there is a lack of safety and security in the rapid development process. So, the risk

involved in different technologies in vaccine development has to be analyzed. Quentin Hass et al, (2021) [28] carried out this analysis using the AI-based search engine Risklick [26, 27]. They consulted research publications since 2000 for collecting data. The logic was developed to control the manufacture and distribution process of the vaccine. This management is important during the time of the pandemic. Mass vaccination around the world could be affected by a small variation in the materials used for vaccine production. The risk score was calculated using an AI driven interface which could predict the crisis faced by the companies developing the vaccines. One year later they compared the result of their work with the contemporary situation. They found that AI-based results predicted correctly the shortage of production and distribution of the RNA vaccine in the future. In a review article by Quoc-viet Pham et al. (2020) [29], the authors explained how AI can be a useful tool for the outbreak prediction of the pandemic and detection of the corona virus. AI can use the data set provided by governments, healthcare organizations, clinical laboratories, and patients for predicting safe and effective drugs/vaccines. Big Data when combined with AI can help healthcare personnel in early diagnosis, and prediction of the treatment result. It can help the governments to predict the outbreak in the future. But the reliability and accuracy of prediction depends on the algorithm which should be optimized properly. Data analytic tools from Oracle cloud computing combine Big Data and AI for vaccine design. Gunjan Arora et al. (2021) [14] reviewed the application of AI in predicting global infection threats. AI can screen patients by using screening tools. These tools can analyze radiological reports like Chest X-rays and CT scans to identify covid patients. ML and other AI techniques can detect the impact of genetic mutation on drug targets. Vaxign, an ML-based reverse vaccinology tool has been able to predict targets that can be used to design COVID-19 vaccines. Alyasseri, Z, et al. (2022) [1] in their article reviewed the role of deep learning and machine learning techniques to interpret the outbreak of the pandemic. They concluded that among ML algorithms SVM is widely used for COVID-19 diagnosis. Similarly, CNN, a DL mechanism is used for outbreak prediction. Arash Keshavarzi Arshadi et al., (2020) [2] studied some viral targets. They found that the main protease and the viral spike protein are the most favorite choices for vaccine development. They formed datasets titled “CoronaDBAI” that can be used for further research applications. Wang et al. (2021) [41,43] reviewed 78 articles discussing the prospect of AI in controlling the COVID-19 pandemic. They observed that AI has a high success rate in pandemic prediction and diagnosis of the disease. AI can be very helpful to the healthcare system during the pandemic. AI can improve the healthcare system during the COVID-19 pandemic. Some recent articles show the challenges and impact made by AI to identify and reuse known drugs available in the market [35]. Rishi R. Gupta (2022) [30] has discussed the use of AI and ML in Bioinformatics. Tania Cova et al (2022) [39] have proposed the use of Artificial Intelligence and Quantum Computing as the next Pharma Disrupters. “Deep Mind”, an AI-based algorithm that predicts the structures of COVID-19 viral proteins useful for drug design. “Kaggle

Platform", used AI to share and analysis of corona virus literature. "Open Research Dataset Challenge by US Government and partner organizations make 29000 research articles available for research. "Bluedot" is an AI-based Early Warning System by WHO that helps in giving early warning by surveillance. John Hopkins University &the Paris-based Organization for Economic Co-operation and Development (oecd.ai) have made an AI-based interactive dashboard to track the spread of the virus. These initiatives will help scientists design effective drugs and vaccines against the highly mutating COVID-19 virus.

ACKNOWLEDGEMENT

The authors are thankful to the University Grant Commission, Government of India for funding.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

Authors declare that there is no conflict of interest.

REFERENCES

1. Alyasseri Z A A, Al-Betar M A, Doush IA, Awadallah MA, Abasi AK, Makhadmeh SN et. al. Review on COVID-19 diagnosis models based on machine learning and deep learning approaches. Expert Systems. 2022;39(3).
2. Arshadi Arash Keshavarzi, Webb Julia, Salem Milad, Cruz Emmanuel, Thomson Stacie Calad, Ghadirian Niloofar et. al. Artificial Intelligence for COVID-19 Drug Discovery and Vaccine Development. Frontiers in Artificial Intelligence. August 2020;3: 65.
3. Breiman L. Random forests. Machine Learning. 2001; 45(1): 5–32.
4. Beck BR, Shin B, Choi Y, Park S, Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Computational and Structural Biotechnology Journal. 2020; 18:784–790.
5. Batool M, Ahmad B, Choi S. A structure-based drug discovery paradigm. International Journal of Molecular Sciences.2019; 20(11):2783.
6. Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA), Effect of shape on binding of steroids to carrier proteins. Journal of the American Chemical Society. 1988;110(18):5959–5967.
7. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995; 20(3):273–297.

8. Dai L, Gao G F. Viral targets for vaccines against COVID-19. Nature Reviews Immunology. 2020; 21: 73–82.
9. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. BMC Biology. 2011;9(1):1–9.
10. Dixon SL, Smondyrev AM, Rao SN. PHASE: a novel approach to pharmacophore modeling and 3D data- base searching. Chemical Biology & Drug Design. 2006;67(5):370–372.
11. Das S, Sarmah S, Lyndem S, and Singha Roy A. An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study. Journal of Biomolecular Structure and Dynamics. 2021; 39(9):3347-3357.
12. Elfiky AA. SARS-CoV-2 RNA dependent RNA polymerase (RdRp) targeting: an in-silico perspective. Journal of Biomolecular Structure and Dynamics. 2021;39(9):3204–3212.
13. Goyal B, Goyal D. Targeting the dimerization of the main protease of coronaviruses: A potential broad-spectrum therapeutic strategy. ACS Combinatorial Science.2020; 22(6): 297–305.
14. Arora Gunjan, Joshi Jayadev, Mandal Rahul Shubhra, Shrivastava Nitisha, Virmani Richa, Sethi Tavpritesh. Review: Artificial Intelligence in Surveillance, Diagnosis, Drug Discovery and Vaccine Development against COVID-19. Pathogens. 2021; 10:1048.
15. Huang SY, Zou X. Advances and challenges in protein-ligand docking. International Journal of Molecular Sciences. 2010; 11(8): 3016–3034.
16. Kalyaanamoorthy S, Chen YPP. Structure-based drug design to augment hit discovery. Drug Discovery Today. 2011;16(17–18):831–839.
17. Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. Journal of Medicinal Chemistry. 1994; 37(24):4130– 4146.
18. Li H, Sutter J, Hoffmann R. Hypo Gen: an automated system for generating 3D predictive pharmacophore models. Pharmacophore Perception, Development, and Use in Drug Design. 2000; 2:171.
19. Lianpan Dai, George F. Gao. Nature Reviews |Immunology. February 2021; 21:73
20. Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-based virtual screening for drug discovery: principles applications and recent advances. Current Topics in Medicinal Chemistry. 2014;14(16):1923–1938.
21. Melo-Filho CC, Braga RC, Andrade CH. 3D-QSAR approaches in drug design: perspectives to generate reliable CoMFA models. Current Computer-Aided Drug Design.2014;10 (2):148–159.
22. Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V. Deep neural nets as a method for quantitative-structure–activity relationships. Journal of Chemical Information and Modeling. 2015;55(2):263–274.
23. Zhu Na, Zhang Dingyu, Wang Wenling, Li Xingwang, Yang Bo, Song Jingdong. China Novel

Coronavirus Investigating and Research Team; A novel coronavirus from patients with pneumonia in China, 2019,2020. New England Journal of Medicine. 382(8): 727–733.

24. Patel H, Noolvi M, Sharma Poonam, Jaiswal V , Bansal S, Lohan Sandeep et. al. Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. Medicinal Chemistry Research. 2014; 23(12): 4991–5007.
25. Patel L, Shukla T, Huang X, Ussery DW, Wang S. Machine learning methods in drug discovery. Molecules.2020; 25(22):5277.
26. Pillaiyar T, Meenakshisundaram S, Manickam M. Recent discovery and development of inhibitors targeting corona viruses. Drug Discovery Today.2020;25(4):668–688.
27. Haas Quentin, Borisov Nikolay, Vicente Alvarez David, Ferdowsi Sohrab, Mayenn Leonhard von, Teodoro Douglas et. al. Vaccine Development in the Time of COVID-19: The Relevance of the Risklick AI to Assist in Risk Assessment and Optimize Performance. Front. Digit. Health. 02 November 2021.
28. Haas Quentin, Vicente Alvarez David, Borisov Nikolay, Ferdowsi Sohrab , Mayenn Leonhard von, Trelle Sven. Utilizing artificial intelligence to manage COVID-19 scientific evidence torrent with Risklick AI: a critical tool for pharmacology and therapy development. Pharmacology. 2021; 106:244–53.
29. Quoc-Viet Pham, Dinh C. Nguyen, Thien Huynh-The, Won-Joo Hwang, Pubudu N. Pathirana. Artificial Intelligence (AI) and Big Data for Corona virus (COVID-19) Pandemic: A Survey on the State-of-the-Arts. IEEE ACCESS. 2020; 4.
30. Gupta Rishi R. Application of Artificial Intelligence and Machine Learning in Drug Discovery. Methods Mol Biol. 2022; 2390:113-124.
31. Joshi Rakesh S, Jagdale Shounak S, Bansode Sneha B, Shiva Shankar S, Tellis Meenakshi B, Pandya Vaibhav Kumar et. al. Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease. Journal of Biomolecular Structure and Dynamics. 2020;39(9):1–16.
32. Sammut C, Webb GI. Naïve Bayes BT. Encyclopedia of Machine Learning, Springer Science & Business Media.2010.
33. Selvaraj C, Dinesh DC, Panwar U, Abhirami R, Boura E, Singh SK. Structure-based virtual screening and molecular dynamics simulation of SARS-CoV-2 guanine- N7 methyltransferase (nsp 14) for identifying antiviral inhibitors against COVID-19. Journal of Biomolecular Structure and Dynamics, 2021; 40(24):13965-1397.
34. Sobhia ME, Ghosh K, Sivangula S, Kumar GS. Deciphering COVID-19 Enigma by Targeting SARS-COV-2 Main Protease Using IN-SILICO Approaches. IJPSR .2021; 12(6) :3104-3119.
35. Stefania Monteleone, Tahsin Kellici F, Southey Michelle, Bodkin Michael J, Heifetz Alexander. Fighting COVID 19 with Artificial Intelligence. Methods Mol Biol. 2022; 2390:103-112.

36. Khan SA, Zia K, Ashraf S, Uddin R, Ul-Haq Z. Identification of chymotrypsin-like protease inhibitors of SARS- CoV-2 via integrated computational approach. Journal of Biomolecular Structure and Dynamics.2020; 39(7):2607–2616.
37. Skariyachan Sinosh, Gopal Dharshini, Chakrabarti Shweta, Kempanna Priya, Uttarkar Akshay, Muddebihalkar Aditi G. Structural and molecular basis of the interaction mechanism of selected drugs towards multiple targets of SARS-CoV-2 by molecular docking and dynamic simulation studies-deciphering the scope of repurposed drugs. Computers in Biology and Medicine. 2020; 126:104054.
38. Teodoro D, Ferdowsi S, Borissov N, Kashani E, Vicente Alvarez D, Copara J et. al. Information Retrieval in an Infodemic: The Case of COVID-19 Publications. J Med Internet Res. 2021.
39. Cova Tânia, Vitorino Carla, Ferreira Márcio, Nunes Sandra, Rondon-Villarreal Paola, Pais Alberto. Artificial Intelligence and Quantum Computing as the next Pharma Disrupters. Artificial Intelligence in Drug Design. 2021; 321–347.
40. Van Drie JH. Generation of three-dimensional pharmacophore models. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2013; 3(5):449–464.
41. Wang L, Ding J, Pan L, Cao D, Jiang H, Ding X. Artificial intelligence facilitates drug design in the big data era. Chemometrics and Intelligent Laboratory Systems. 2019;194:103850.
42. Wahedi HM, Ahmad S, Abbasi SW. Stilbene-based natural compounds as promising drug candidates against COVID-19. Journal of Biomolecular Structure and Dynamics. 2021;39(9):3225–3234.
43. Wang L, Zhang Y, Wang D, Tong X, Liu T, Zhang S et. al. Artificial Intelligence for COVID-19: A Systematic Review. Front. Med. 2021; 8:704256.
44. Dai Wenhao, Zhang Bing, Jiang Xia-Ming, Su Haixia, Li Jian, Zhao Yao et. al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science.2020; 368(6497): 1331–1335.