International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 12 Issue 8 ǁ August 2024 ǁ PP. 41-51

Application of Artificial Intelligence (AI) in the Analysis and Computer Aided Drug Design (CADD) of Lung Cancer

Anil Kumar Mishra[1], Janardan Prasad Pandey*[2]
[1]*Department of Physics*
M G P G College, Gorakhpur U.P.
[2]*M L K P G College, Balrampur U.P.*

Abstract
The rapid advancements in Artificial Intelligence (AI) have significantly impacted various sectors, including healthcare. This review explores the application of AI in the analysis and Computer-Aided Drug Design (CADD) specifically targeting lung cancer, a disease with high mortality due to late diagnosis and limited treatment options. AI, through sophisticated algorithms, enhances early detection and accurate diagnosis of lung cancer by identifying cancerous cells and mutations at initial stages. The integration of AI in CADD streamlines the drug development process, facilitating the design and optimization of potential drugs. Techniques such as structure-based drug design (SBDD) and ligand-based drug design (LBDD) utilize AI to predict molecular interactions and optimize therapeutic compounds. The review also highlights the role of physics-based methods in CADD, providing a detailed understanding of molecular dynamics, quantum mechanics, and pharmacophore modeling. By leveraging AI and CADD, researchers can accelerate the discovery of effective treatments, offering new hope in the fight against lung cancer. However, the ethical and responsible application of AI in healthcare remains crucial to ensure patient safety and efficacy of treatments.

Date of Submission: 01-08-2024 Date of acceptance: 09-08-2024

I. INTRODUCTION

Artificial intellect (AI) is the imitation of human intellect in computers that have been designed with human-like thought and learning processes. Artificial Intelligence (AI) aims to develop systems that are capable of activities that typically require human intelligence, like speech recognition, visual perception, language processing, and decision-making. The development of AI involves creating algorithms and models that enable machines to perform tasks without explicit programming, thus allowing them to learn and improve from experience. AI strives to be equivalent to human intelligence in that it must be able to comprehend, learn, and apply knowledge to a wide range of tasks. AI already has a strong effect on various sectors like healthcare, education, finance, entertainment etc. In recent times, technology, including artificial intelligence and telemedicine, has been increasingly integrated into healthcare systems, enhancing accessibility, efficiency, and the quality of care. The continued evolution of healthcare systems is crucial for meeting the diverse and evolving health needs of populations worldwide.

Quick progress of today's technology has resulted in medication and cure of several fatal diseases including cancer, which has the 2nd highest global mortality rate in all over the world. Recent studies have shown that lung cancer is the leading cause of cancer related deaths and overall, it is the second most common. This high mortality and low survival rate are because of late diagnosis and limited treatment methods and medication. Mostly patients are diagnosed in the advanced stage and before that they are asymptomatic and hard to diagnose which leads to the high mortality rate. So, in order to reduce the mortality rate of lung cancer, early diagnosis and accurate treatment of lung cancer is very important. Traditional approaches for this are very limited and prone to human error while artificial intelligence based on algorithms can identify cancerous cells and mutations in the early stages of cancer and are the need of the hour. AI holds a very significant part in the diagnosis and targeted drug design for cancer. Using computational methods AI can find the correct computations of a drug, design it and also analyse the performance of a biological target. AI uses big training data set to predict the future outcomes of a particular drug which is known to it previously, but while dealing with unknown molecules PHYSICS-BASED METHODS are used to help AI predict the molecular properties of a molecule while binding to a protein, without the need of existing data. Now let us discuss these in detail in the following sections.

II. A DETAILED STUDY ON AI

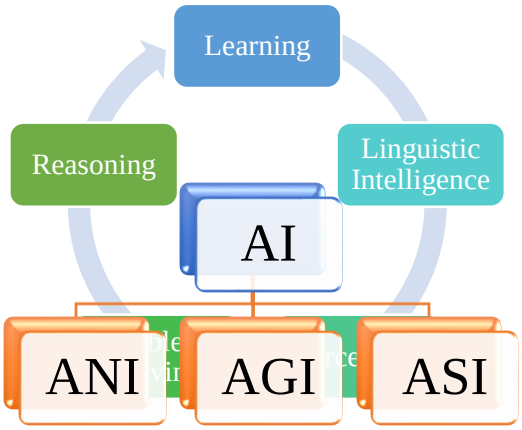
Artificial Intelligence is a man-made interface which acts just like human intelligence and hence the name artificial. It has been a part of our daily lives since years. Starting from google to Netflix to Spotify to any other kind of platform, they all rely on AI to make personalised suggestions based on algorithms. Many a times we don't even realise that we are in fact using the AI or its algorithm to cater our needs. Computers that have been taught to think and learn like people are simulating human intelligence. They can do tasks like speech recognition, visual perception, decision-making, and language understanding without the assistance of people. AI works by using the large data set that is being provided to it. It analyses the vast data, searches for patterns and co relations and similarities as well as dissimilarities and uses these patterns to predict the future outcomes based on the provided data set. Unlike humans it can process the large amount of data set, that is already available easily in today's world due to all kinds of social media platforms and advancements of technologies, using different methods and then it can make predictions for future.

2.1 PARTS OF INTELLIGENCE:

1. Learning
2. Reasoning
3. Problem Solving
4. Perception
5. Linguistic Intelligence

2.1.1 Learning: Just like a human being learns to do things in a certain way through experienced and trial and error and gradually learns, AI also has the capacity to learn from its experience with the data and improves itself from the feedback. It is the most crucial step in any AI program. Without explicit programming of a human, the AI learns through sorting out the data, labelling and finding similar patterns.it positive and negative feedbacks and help AI in its learning. It memorizes its past experience and uses it for similar problems in future.

2.1.2 Reasoning: The second most important component of AI is the ability to judge and make decisions based on the inferences drawn. Through inductive and deductive reasoning, it draws conclusions from a situation and without any human interference. It has allowed the software to generate consistent results using various logical rules and algorithms.



2.1.3 Problem Solving: It is similar to reasoning and is the most pivotal component for the development of any AI system. As the name suggests, it is the ability of the system to solve a particular problem. It is the process of selecting the most suitable route to look for the best possible solution or the desired goal of a problem. The system takes in the data, manipulates it to create a solution to solve the problem in hand.

2.1.4 Perception: In humans, sensory organs are responsible for perception., likewise in AI perception refers to analysing, interpreting a picture or any physical object in a given environment, video recognition etc. in AI, the data acquired by the sensors are put together in a meaningful manner for the system to perceive and understand.

2.1.5 Linguistic Intelligence: It is the ability of one to speak, understand, use, comprehend and write the verbal and written language. It the final component which makes up the artificial intelligence. It is important for the ease of one-to-one conversation between the system and the operator. Through this ability the system understands the language used by the user and not only the machine language and also gives output in the same.

2.2 TYPES OF AI

1. Artificial Narrow Intelligence (ANI)
2. Artificial General Intelligence (AGI)
3. Artificial Super Intelligence (ASI)

Based on the capabilities and functions AI is categorised into these three types.

1. **Artificial Narrow Intelligence (Ani)**: The only kind of artificial intelligence that has been completely developed and is currently in use on the market is artificial narrow intelligence, or weak AI. This the only type which truly exists in our era. This type of AI specializes in only one type of solution and is designed for a single task which it completes even in a better way than humans. It does not possess broad cognitive abilities like humans. It is limited to a specific domain, excels in that particular goal.
2. **Artificial General Intelligence (Agi):** It is also known as strong AI. It has the intelligence comparable to human intelligence. It is only a thing of research for now. It has not been yet fully realized till date. It can

perform any generalized task that a human can. It can analyse, comprehend, learn and understand a wide range of task with a high degree of versatility and adaptability, much like a human being. It has the ability to use its intelligence to solve any given problem in hand. Researchers and scientists are still researching on this AGI. Given that we still do not have the full knowledge about how the brain function, making an artificial replica of such a complex brain structure with neural activities is a difficult and matter. Scientists are hopeful for it to be a reality in the near future.

3. Artificial Super Intelligence (Asi): Artificial super intelligence is only a hypothetical thing, that we often see in science fiction movies and scenes. It's the kind of AI that not merely matches, but exceeds human intelligence. ASI can self-analyse and become self-aware, build relationships, be creative and surpass humans in every way possible. It has emotions, beliefs, desires and needs of its own just like humans. It can decide to be good or evil for itself. Once we achieve AGI, ASI would not be a huge change. When AI would become comparable to humans, the next logical thing would be for themselves to try to be better than humans which would give rise to ASI. It still has a long way to go. Whatever would happen afterwards is purely Imaginary.

2.3 DOMAINS OF AI:

Broadly there are five main domains of AI. As technology continues to advance, AI diversifies and has a wide variety of applications and domains. It encompasses various specialized areas where AI is being developed to carry out specific tasks and applications. Some major domains of AI are as followed:

1. MACHINE LEARNING
2. DATA SCIENCE
3. COMPUTER VISION
4. NATURAL LANGUAGE PROCESSING

2.3.1 MACHINE LEARNING:

It is the most common form of ANI that is being widely used worldwide. It is a subset of AI that uses data and past experience to gain intelligence. It is the most common form of AI that is being used in the world today and predominantly it uses ANI. In the last few years large volumes of data are being generated and machine learning makes use of this data, processes it and further makes predictions based on the trained algorithms and statistical methods. There are various types of machine learning which are listed below.

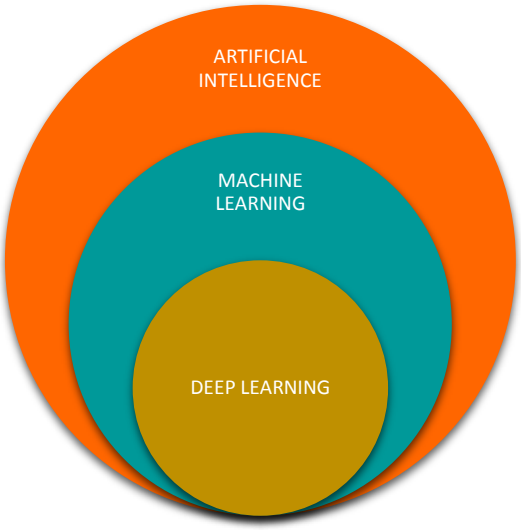
A. Supervised Learning: In supervised training/learning, the algorithm develops a model based on the training of a labelled dataset. In this type of learning, both the inputs and the corresponding outputs are provided to the model. The model then maps the inputs to their outputs and makes predictions based on its experience with previous dataset. It practices with the training data set and makes future predictions. The output is then evaluated and the error is minimized using feedbacks and gradually the model gets developed. It cannot work with new or unlabelled data.

B. Unsupervised Learning: In this type of learning, the machine is provided with unlabelled data and it tries to find patterns, structures and relations in those data by using various methods without any strict guidance. The algorithm processes the data itself and divides into groups in its own way without any prior knowledge. The model develops and learns independently from the data set. It can work with new and unknown data and be able to provide outcome. The algorithm looks for underlying patterns and finds new relationships among those data. Through these methods it trains and develops a model.

C. Semi-Supervised Learning: In a big data set, there will always be some data which are labelled and some data which are not. To train in this set of data, semi-supervised learning is used. It stands in between supervised and non-supervised learning. it mostly uses unlabelled data but also has some labelled data to show that some type of classification does exist. Initially it trains with the limited labelled data in a similar manner as supervised learning and keeps on gaining experience and develops the model until the errors are minimized and gets correct result. Then the algorithm tries out the unlabelled data and uses it to find patterns, similarities, clusters and so on and keeps on developing until it gets a stable result in a similar manner as unsupervised training but with a slight difference, that it already has some ideas about the data set from its previous experience with labelled data. It is very much useful when we deal with a large size of unlabelled data set.

D. Reinforcement Learning: In this type of machine learning, the algorithm decides for itself what to do with the input. Based on its decision, a positive or negative reward is awarded to it. From the reward or penalty, it gains the experience of which is correct or which is wrong and it further modifies its behaviours. This type of learning is mostly used in robotics. This learning is basically the art of decision making and interpreting. The model will be given a task to do and there will be no training dataset. So, it is forced to take a decision and the reinforcement agent reviews its decision and the path it took to solve it and based on the behaviour reward is given. After receiving the reward, the model determines if or not the choice made was correct, neutral or wrong.

The model keeps on developing on a trial-and-error basis. And gradually finds its best suited path in the given environment. This is most suited for automated machines to learn by themselves without any human guidance.



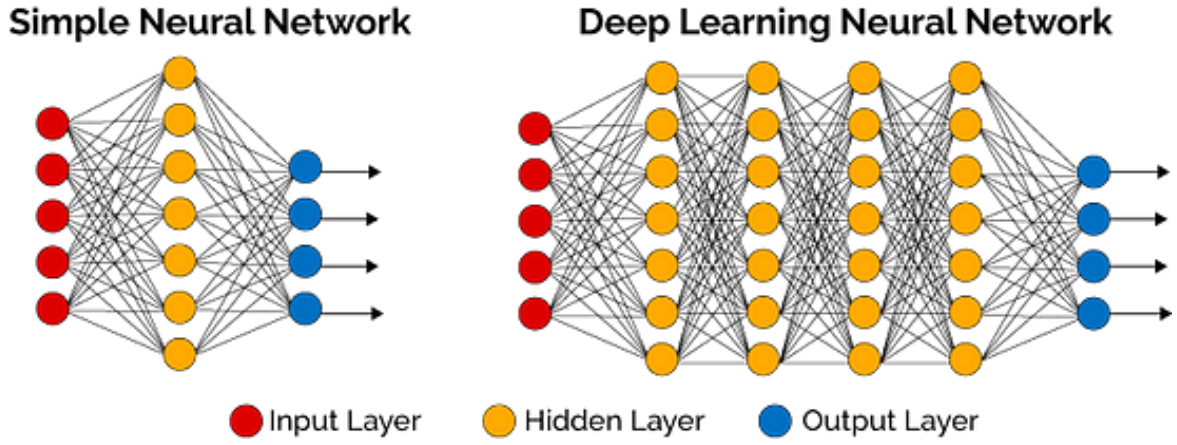
2.3.1.1 DEEP LEARNING:

Deep learning is frequently mistaken for machine learning, but they in fact are very different from each other. Machine learning, which is a subset of artificial intelligence itself, includes deep learning as well. Deep learning is centred on creating artificial neural networks that function similarly to human brain networks in order to tackle challenging issues. It has not yet been fully realised. One of the super computers in the world took 40 minutes to stimulate a 1 sec of neural activity similar to human brain. This alone shows we have a long way to go yet. The growth of deep learning will lead to the realization of AGI. It uses artificial neural networks to identify the hidden patterns and similar points that exist in the data set unlike the machine learning which uses its algorithm to find it. 'Deep' in the term deep learning describes the depth of artificial neural network layers that exist between the input and the output. It uses a vast volume of dataset than that of the machine learning to train itself efficiently.it is relatively more complex and takes more time to train the model.

➢ **Artificial Neural Networks (Ann):** Artificial neural network is an imitation of human biological neural networks. It is a computational structure influenced by human brain. Just like un human neural system there exist neurons, nodes, axons, in ANN also similar basic structures exist. Artificial neurons are called as units. These units are arranged in different layers which in turn constitute the whole ANN system. The more complex the ANN is, the more units are present in it, probably in millions of numbers. These layers identify the patterns hidden in the dataset. In an ANN there are 3 types of layers that exist within. An input layer, an output layer and a hidden layer. The input neurons behave as dendrites and receives the initial data or information or features which it needs to analyse or learn and represent it to the input layer. The hidden layer lies in between the input and output layer. These layers change the input data into a different form which is more convenient for the output layer and helps in the learning process. The output neurons act similar to that of an axon and provides an output for the provided input data. Mostly these units in ANN are interconnected in a very complex manner. Data gets transferred from one layer to another layer and the ANN keeps on learning and developing.

TYPES OF ANN:

➢ **Feedforward Neural Network (FNN):** In this type of neural network, information moves from the input layer towards the output layer in a single path and in a single direction. It is one of the most basic types of neural network to exist. Here the hidden layers may or may not even exist. The data simply enters through the input layer and then exits through the output layer.



➢ **Convolutional Neural Network (CNN):** This type of neural network is usually specialized for image or speech processing or any grid like data. It has several convolutional layers which performs an operation on the input. The resultant output data is then passed to another layer. These layers apply some filters to the dataset and

then gives the final output. These networks are very good at handling large volume of data and can detect underlying patterns and features in images, videos, audios, etc.

- **Recurrent Neural Network (RNN):** This type of neural network forms a cyclic loop. First few layers in this are mostly present in a similar manner to that of an FNN. After the output of the first few layers are completed, the RNN starts. This network does not give away the whole output, instead it remembers a part of the output and feed this back to the input to be used. This helps in a better prediction. Each unit of this RNN gives the output and remembers some data from its previous step which serves as a memory cell for that unit, making the learning more efficient.

2.3.2 DATA SCIENCE:

Data science is the study of colossal data and drawing conclusions for our needs. This data is converted into human readable format using AI. For the functioning of any AI system data is the most necessary and basic component which we absolutely need. We use data mining processes to find meaningful patterns and groups from the huge data set. Data mining includes visualization of data, statistical and analytical methods, machine and deep learning and neural networks. There are many types of data available to us in the world. We can categorise them into structured, unstructured or semi-structured data. Based on the types of data used in any AI system, data can be divided into either 'training dataset' or testing 'dataset'. The data used to teach an AI algorithm how to complete a task is referred to as training data and the data used to evaluate the accuracy and functionality of the created AI model is referred to as testing data.

I. **Problem Scoping:** First, we try to identify the problem for which we need the solution for. After identifying the problem, we can proceed to find the solution. For the problem identification, we use the 5 W' method i.e., asking the question with who, what, why, where and when. Answering these 5 W's we can identify the question and the solution that we will need.

II. **Data Acquisition**: In this step we try gather all sort of relevant data and information for our problem to be used as training or testing data for the model. In this phase we collect all type of data from various available sources and create a database.

III. **Data Exploration:** After creating and compiling the database, we analyse the data and draw meaningful results and conclusions from it. In this step we identify important patterns and trends for a better understanding of the data and to get insights. We basically visualise the data overall and hence this step is also called data visualization. In forms of different charts, bar graphs, pie graphs we can do this visualization.

IV. **Modelling:** In this stage we train our models using the collected data set. We apply various modelling techniques to train our model and make future predictions through this. The model uses training data set and keeps on developing using these modelling techniques.

V. **Evaluation:** This is the last stage of any AI project. In this stage the developed AI model is tested for its accuracy and performance using testing data set. After the test it is evaluated if or not it can be used for real life use. The final perfect model is then deployed for real life use after analysing the feedback and making necessary improvements.

2.3.3 COMPUTER VISION:

Computer vision is another domain of AI. It is the ability of the computers to recognise, process and presume the images and videos. It imitates human vision and contributes towards an automated system which can understand and draw meaningful conclusions from visual data and takes actions accordingly. It uses machine learning and different modelling techniques to teach itself from the visual data and to further recognise an image. It has 5 stages as shown. These processes make the computer perceive the visual content and behave accordingly.

2.3.4 NATURAL LANGUAGE PROCESSING:

This domain of AI is responsible for its linguistic intelligence. It enables the system to comprehend and process human language in a proper and useful way. Using this, the AI model is able to interact with humans using natural languages. Humans can directly talk one to one with the AI model. NLP processes the natural unstructured human data into a machine-readable structured data. NLP is used in a variety of areas and even in our daily lives. It is used for machine translation as well as to understand underlying emotional meaning of the text. Most commonly it is used for virtual assistance. Processes involved in this are as followed.

3. USE OF AI IN LUNGS CANCER:

3.1 For Quality Scan: A minimal quality level is necessary for AI or humans to be able to diagnose a CT scan which is the basic necessary thing to do for a lung cancer detection. It is especially crucial to maintain the lowest feasible radiation exposure during the screening process. An average effective dose of 1.5 mSv is already sufficient to get excellent quality images because of the strong contrast between the lung tissue and air. The majority of lung cancer screening trials, including the NLST, employed these small dose CT scans. Thanks to

technological advancements, iterative reconstruction techniques have been applied in clinical settings since 2009. This method makes several revisions to each rebuilt image to eliminate artifacts and enhance overall image quality. Ultra-low-dose CT scans were also made possible by this advancement; they have an average radiation dose of roughly 0.5 mSv, which is comparable to that of X-rays when it comes to chest scanning. Recently, DL techniques have also been applied to enhance radiation dosage and rebuilding time. According to a pilot investigation, all nodules larger than 2 mm that were discernible on typical low-dose CT scans were likewise seen on ultra-low-dose imaging.

3.2 For Detection of Nodule: Pulmonary nodules, or opacities in the lung parenchyma that are not thought to be a natural component of the anatomy, are the classic visual presentation of lung cancer on CT scans. Finding every pulmonary nodule is the first step in the process of diagnosing lung cancer. It is common knowledge that not all nodules are detected by radiologists. Recognize and characterize pulmonary nodules which is something specific in a picture packed with airways and vessels, is a challenging assignment for humans, particularly when it is unclear how many nodules are present in the first place. The AI algorithm has method achieved a sensitivity of 97.2% in nodule detection at the cost of an average of 1 false positive per scan.

3.3 Classification And Measurement of Nodule: Guidelines divide nodules into categories according to their size and type, which are the two key factors that determine their risk of malignancy. Ciompi created an artificial intelligence program to distinguish between six different forms of nodules. The algorithm's goal was to automatically categorize nodule types that are clinically significant. The AI algorithm performed similarly to an independent human expert and its ability to automatically classify pulmonary nodules in lung cancer screening was very much reliable. The greatest CT indicators of cancer are unquestionably bigger nodules and nodular development. The largest and orthogonal diameters in the transverse plane are individually measured to estimate size. Although volumetric segmentation techniques tend to be more repeatable, majority lung cancer screening trials did not employ them, and they are not widely accessible. The produced three-dimensional segment may also be automatically used to calculate the diameter. According to a recent study, when included into a multivariable logistic regression model, the mean diameter obtained from a CAD is as predictive of malignancy as the volume produced from the CAD.

3.4 Prediction Of Malignancy: The main objective of CT lung cancer screening is to diagnose lung cancer in an individual. Malignancy probability is calculated in the screening procedure subsequent to the identification and CT nodule characterization. The Brock model is the most well-known quantitative risk model for assessing the probability of nodule malignancy. On further additional screening and clinical datasets, it has performed well. Numerous factors related to patient demographics, nodule size, kind, and shape are incorporated into the Brock model. The most significant indicator of malignancy on CT scans is nodule development, which cannot be determined from just one image. A neural network that was created recently to evaluate the risk of lung cancer using follow-up CT scans performed well on a dataset that was independent of the study.

According to the preceding section, cutting-edge AI algorithms for lung cancer CT diagnosis have possibly reached performance comparable to that of radiologists. However, these investigations did not take into account a partnership between a human and a machine; instead, they merely examined individual performances. The computer program may find some jobs simpler than those that the radiologist finds more challenging, and vice versa. For instance, it is well-established that subsolid nodules are more likely to be overlooked by radiologists because they have a lower contrast from the lung parenchyma; on the other hand, extremely irregular nodules can be difficult for the AI to identify since they are uncommon in the data set used for training.

4. COMPUTER AIDED DRUG DESIGN (CADD) FOR LUNG CANCER:

Designing a drug is the creative procedure for developing novel drugs using biological target information. A potential drug research lead is found and developed using a combination of computer technologies known as computer-aided drug design (CADD). Molecular modelling, computational chemistry, and molecular design are all included in CADD. Access to vast amounts of data does not ensure the acquisition of useful prediction models in the big data world of today. High volume, complex, and sparse data sources need to be carefully addressed in order to anticipate treatment efficacy and adverse effects.

4.1 COMPUTER-AIDED DRUG DESIGN (CADD):

In the process of finding new drugs, CADD filters out huge compound libraries into smaller groups of likely active compounds. This allows main compounds to be optimized by enhancing their biological characteristics and creating chemotypes from nucleating sites by integrating fragments with tailored functions. Screening hits comprise both molecules that bind to the target specifically and a larger number of nonspecific substances that need to be filtered away by a sorting process. The compounds are sorted across a wide range of classes, computational techniques have been applied to categorized on the basis of structural similarity. The selection would therefore depend on cluster, potency, and other elements like ligand efficiency. The growing use of various computational techniques in drug development has made it possible to handle data related to many

compounds more effectively when they are evaluated for matches against target molecules or proteins. The discovery of lead molecules from databases has been made possible by computational methods, which also aid in defining and elaborating the degree of interaction between ligands and targets. With a significant cost and time advantage, the CADD technique has been important in the discovery and optimization of possible lead compounds. It has been used at several phases of the drug development process, including target validation, interaction analysis, molecular design, and identification of potential targets. CADD employs two distinct methodologies, contingent upon the accessibility of either ligands or three-dimensional protein structures. These are generally known as structure-based drug design (SBDD) and ligand-based drug design (LBDD). They can be applied individually or can be integrated to find accurately the targeted molecule.

4.1.1 STRUCTUREBASED DRUG DESIGN (SBDD):

The knowledge of the three-dimensional structures of the therapeutic target protein and the study of the binding site cavity are the fundamental ideas of structure-based drug design. The disclosure of the 3D structures of numerous biological compounds has resulted in a new age of SBDD in discovery and design of drugs. In the pharmaceutical business, SBDD has gained attention as a potential method for producing and refining ligands. The fundamental phases of SBDD include target preparation, binding site identification, molecular docking, molecular dynamics and virtual screening.

1. Molecular Docking (Md): The method known as "molecular docking" is used to ascertain the positions and configurations of ligand molecules within a macromolecular target's binding site. After creating poses with search algorithms, scoring techniques are used to rank them. Molecular recognitions such as those between an enzyme and its substrate, a drug and a protein, a nucleic acid and a drug, and a protein and its associated protein interact to power numerous biological activities, including signal transduction, cell regulation, and other macromolecular assemblies. Two essential elements of a protein-ligand docking technique are sampling and scoring. Two aspects of sampling, related to forming potential ligand binding orientations close to the binding site of a protein, are ligand sampling and protein flexibility. Applying physical or empirical functions, scoring forecasts binding tightness for specific ligand orientations and configurations.

2. Virtual Screening (Vs): Virtual screening is the term for the techniques used for choosing required compounds from chemical databases. It can be thought of as the computerized version of high-throughput screening and other experimental biological evaluation techniques. Utilizing the compound libraries which are huge and chemically diverse, for computational and biological research is a widely used approach in medication development. In computational drug discovery, virtual screening is the most typical situation. VS can be viewed as a technique for ranking molecules for experimental study, much as computational mutagenesis. While VS has been thoroughly investigated using SB and ML techniques separately, its combination, either fully combined or sequentially, allows for the utilization of the greatest amount of accessible data and, predictably, yields more accurate findings. VS can be separated into two groups: (I) Structure-based Virtual Screening (SBVS) and (II) Ligand-based Virtual Screening (LBVS)

4.1.2 LIGAND BASED DRUG DESIGN (LBDD):

Drug design based on ligands is utilized when three-dimensional information about the receptor is unavailable. The method depends on an understanding of the molecules which attach to our desired biological target. By using an antibiotic (ligands)which is already known, as a target, LBDD techniques create a structure-activity relationship (SAR) among the physiochemical characteristics of the target and the antibiotic behaviours. This information may be used to drive the creation of novel medications with increased activity or to improve currently available ones. Two approaches for LBDD are Pharmacophore and Quantitative-structure Activity Relationship (QSAR). Molecular similarity techniques use chemical fingerprints of established ligands that bind to targets to find compounds that share a fingerprint while screening molecular libraries. Due to the same binding characteristics of structurally related molecules, ligand similarity search techniques are productive.

1.Quantitative-Structure Activity Relationship (QSAR): A computerized statistical method that describes the perceived range of structural changes brought about by substitution is called a quantitative structural activity relationship, or QSAR. These models show, mathematically, how a ligand's structural characteristics impact the target it binds to, activity response. Molecular characteristics as hydrophobic, electronic, steric, and sub-structural impacts can be utilized to construct QSAR models.

2.Pharmacophore: "An assemblage of electronic and steric characteristics essential for effective supra-molecular interactions with a particular biological target as well as trigger or inhibit its biological response" is the definition of a pharmacophore, as stated by IUPAC. A broad representation of the anatomical characteristics needed for a biological macromolecule to identify the ligand is called a pharmacophore.

4.2 APPLICATION OF AI IN CADD:

AI is primarily employed as a tool to aid patients during operations or other medical procedures., to assess the severity of a patient's condition and predict whether a particular therapy will be effective for them prior to it is administered, and to prevent or address potential problems during treatment. It dictates the usage of particular tools or medications over the course of treatment. In order to improve safety and effectiveness, it creates or calculates new uses for such tools or medications. Large amounts of pre-trained models are often consumed, the data is analysed for relationships and trends, and the patterns are then used to make forecasts. AI is capable of optimizing drug structure design, identifying target and lead molecules, and validating the drug target more quickly. Moreover, it can help with new drug design, interactions between proteins, pharmacological effect, and the three-dimensional shape prediction of a specific protein. Natural Language Processing, Markov Decision Process, Artificial Neural Networks, Support Vector Machines, and heuristics are some basic AI approaches.

PHYSICS APPLICATION IN CADD:

Physics is important to CADD in a number of ways. Physics is extensively included into CADD's techniques and algorithms. CADD's computational techniques are based on fundamental principles of physics. Researchers can simulate and comprehend the intricate processes involved in drug design and optimization through the application of physics, whether it is through the simulation of molecular dynamics, the prediction of quantum-level interactions, or the estimation of thermodynamic properties. The effectiveness and success rate of drug development initiatives are improved by these computational methods. By using physical principles, scientists may anticipate drug-target binding, evaluate the pharmacological characteristics of possible treatment options, and model and study molecular interactions. The drug development process is made much more successful and efficient by these computer methods. Some of the used aspects are:

4.2.1 Molecular Dynamics Simulation: Molecular dynamics simulations, based on physics, are utilized to examine the motion and conduct of molecules across time. Through these simulations, scientists are able to acquire an understanding of the interactions in the atomic level that occur in between a drug candidate and its target protein. The concepts of statistical thermodynamics and classical mechanics offer insights into the flexibility and stability of drug-protein complexes. The foundation of molecular dynamics simulations is classical mechanics, in which the motions of atoms and molecules over a period of time are modelled by using Newton's equations of motion. The flexibility, stability, and structural changes of biomolecules are dynamically shown by MD simulations. By using these simulations, we learn more about the binding mechanisms and possible allosteric effects of drugs by studying how they interact with their target protein.

4.2.2 Quantum Mechanics: Quantum mechanical calculations can provide detailed information on molecular orbitals, electronic states, and energy levels, contributing to the understanding of the chemical properties of drug candidates. Quantum mechanics is concerned with the behaviour of particles at the atomic and subatomic levels. It involves the Schrödinger equation and provides a more accurate description of electron behaviour. QM methods are applied to study the electronic structure of molecules accurately. These methods are crucial for understanding the energetics of molecular interactions, such as bond formation and electron density distribution.

4.2.3 Molecular Docking (MD): Molecular docking estimates the binding affinity between the ligand and the targeted protein using scoring algorithms based on physics. The ideas of force fields, which simulate the interactions between atoms and molecules, are incorporated into these scoring functions. The intermolecular forces are modelled using a variety of force fields that are drawn from physics concepts. These force fields include van der Waals forces, hydrogen bonding, and electrostatic interactions. The binding mode and strength of interactions between a drug candidate and its target are predicted using docking simulations. The selection of possible lead compounds for additional research is aided by this knowledge.

4.2.4 Free Energy Calculations: Based on the laws of thermodynamics, free energy calculations are used to determine the binding free energy of a drug candidate to its target. The computation of free energy relies on the fundamentals of thermodynamics, which include ideas like Gibbs free energy, enthalpy, and entropy. Determining the binding free energy of a ligand to its target by free energy calculations is essential for comprehending the thermodynamic stability of drug-protein complexes. The selection of lead compounds with favourable binding energetics is guided by this knowledge. Generalized Born surface area (MM-GBSA) and molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) are two examples of techniques that use physics-based methods to calculate free energy.

4.2.5 Pharmacophore Modelling: Finding the key components of a pharmacological molecule that contribute to its biological activity is known as pharmacophore modelling. Understanding a molecule's steric and electronic characteristics, which are essential for biological activity, is the foundation of pharmacophore modelling. This method frequently makes use of ideas from molecular physics and quantum mechanics to comprehend the steric and electronic characteristics of functional groups. Pharmacophore models facilitate the identification of essential molecular characteristics needed for ligand-target binding. New compounds are designed or old ones are optimized with the use of this information.

4.2.6 ADMET Predictions: Physical principles are widely used to estimate physicochemical qualities like as solubility, permeability, and metabolic stability, which are important in the prediction of Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET). Physics based computational models aid in the estimation of solubility, permeability, and metabolic stability, among other factors.

4.3 DEVELOPED DRUGS FOR LUNG CANCER USING CADD:

Several drugs for lung cancer have been developed using a combination of traditional drug discovery methods and modern computational techniques, such as CADD. Examples include tyrosine kinase inhibitors like erlotinib and gefitinib, which target specific mutations in lung cancer cells. Some drugs developed for lung cancer, where computational methods like CADD played a role, include:

4.3.1 Erlotinib (Tarceva): An epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor which is used to treat non-small cell lung cancer.

4.3.2 Gefitinib (Iressa): Another EGFR tyrosine kinase inhibitor used in the medication of non-small cell lung cancer.

4.3.3 Afatinib (Gilotrif): It is an irreversible ErbB family blocker which also includes EGF and is used in the treatment of non-small cell lung cancer.

Drug development, such as the development of erlotinib, usually involves a multi-step process in which Computer-Aided Drug Design (CADD) is essential at various phases. Tyrosine kinase inhibitors such as erlotinib are used to treat non-small cell lung cancer (NSCLC). It functions by focusing on and blocking the epidermal growth factor receptor (EGFR), a particular protein. How erlotinib works:

- **Epidermal Growth Factor Receptor (EGFR):** EGFR is a kind of protein which has a very important part in regulating growth of cells and their division. It is often overexpressed or mutated in certain types of cancer, leading to uncontrolled cell proliferation.
- **Tyrosine Kinase Inhibition:** Erlotinib is a member of the tyrosine kinase inhibitor drugs class. (TKIs). EGFR is tyrosine kinase, meaning it is involved in the transfer of phosphate groups to tyrosine residues in proteins, a process that is part of cell signalling pathways.
- **Blocking Signal Transduction:** Erlotinib inhibits the tyrosine kinase activity of EGFR. By doing so, it interferes with the EGFR signalling pathway, which normally transmits signals for cell growth and survival. This inhibition prevents the activation of downstream signalling pathways that promote cell division.
- **Cell Cycle Arrest and Apoptosis:** Interruption of the EGFR signalling pathway leads to cell cycle arrest, preventing cancer cells from progressing through the cell cycle and dividing. Additionally, erlotinib induces apoptosis (programmed cell death) in cancer cells, contributing to the reduction of cancer cell viability.
- **Anti-Angiogenic Effects:** Erlotinib may also have anti-angiogenic effects, meaning it can inhibit the formation of new blood vessels that tumours need for sustained growth.
- **Specific Patient Selection:** Erlotinib is frequently prescribed based on the presence of specific variations or alterations in the EGFR gene in the patient's tumour. Not all patients will benefit equally from EGFR inhibitors, and genetic testing may guide treatment decisions.

4.4 AN OVERVIEW OF THE CADD PROCESS FOR THE DEVELOPMENT OF ERLOTINIB:

- **Target Identification and Validation:** Identify a molecular target associated with the disease, in this case, a relevant tyrosine kinase which is involved in the proliferation and growth of cancer cells. The target's role in the disease is validated through experimental and clinical studies.
- **Virtual Screening:** Use computational techniques to screen large chemical databases for potential compounds that could interact with the identified target. Employ molecular docking simulations to predict the binding affinity and orientation of candidate compounds to the target.
- **Hit Identification and Optimization:** Identify lead compounds from the virtual screening that show promising interactions with the target. Optimize the chemical structure of the lead compounds using CADD methods to enhance binding affinity, selectivity, and pharmacokinetic properties.
- **Pharmacophore Modelling:** Develop a pharmacophore model to understand the essential features of the lead compounds that contribute to their biological activity. Use this model to guide further optimization and design of new compounds.
- **ADMET Prediction:** Prediction of Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) characteristics of the lead compounds to assess their likelihood of success in clinical development. CADD tools can help identify potential issues and guide modifications to improve the drug's overall profile.
- **Preclinical Testing:** Synthesize the optimized compounds and conduct preclinical studies to evaluate their safety, efficacy, and pharmacokinetics. CADD tools may aid in the interpretation of experimental data and guide further iterations in the drug design.

- **Clinical Trials:** If the preclinical studies are successful, the drug candidate progresses to clinical trials involving human subjects. CADD tools may continue to be used for optimizing dosages, predicting potential side effects, and refining the drug's properties.
- **Regulatory Approval and Market Launch:** Successful completion of clinical trials may lead to regulatory approval, allowing the drug to be marketed and prescribed for the treatment of the targeted disease.

Throughout this process, CADD contributes by accelerating the identification of potential drug candidates, guiding optimization efforts, and providing insights into the molecular interactions between the drug and its target. It helps prioritize compounds for synthesis and experimental testing, ultimately contributing to the efficiency of the drug development pipeline.

V. CONCLUSION:

Artificial intelligence (AI) shows great promise for revolutionizing the discipline of oncology in the areas of drug creation and lung cancer treatment. The use of CADD has grown in importance in the search for novel cancer treatments. The methodology integrates computational models and simulations with empirical data to forecast the behaviour of small molecules towards cancer-related biological targets. This facilitates the expeditious discovery of plausible therapeutic options and the refinement of their characteristics to optimize effectiveness and reduce toxicity. A number of computational techniques have been applied to the field of cancer drug discovery. Small molecule activities against specific cancer-related biological targets have been predicted. Multi-target drug designs aimed at multiple biological targets at once have been created and the molecular processes of behaviour of small molecules in cancer have also been predicted. Notwithstanding its achievements in identifying novel cancer treatments, CADD continues to confront obstacles, such as the shortcomings of the predictive models in use today and the requirement for more advanced models capable of explaining the intricate interactions among several biological targets and pathways in cancer. AI's potential applications in medication development and lung cancer treatment are fields that are developing quickly. The application of AI technology could revolutionize our understanding, diagnosis, and treatment of lung cancer, ultimately leading to better patient outcomes and advancements in the oncology sector. Nonetheless, in order to guarantee the ethical and responsible application of AI in healthcare, it is imperative to address issues and concerns.

REFERRENCES

[1]. Chiu HY, Chao HS, Chen YM. Application of Artificial Intelligence in Lung Cancer. Cancers (Basel). 2022 Mar 8;14(6):1370. doi: 10.3390/cancers14061370. PMID: 35326521; PMCID: PMC8946647.

[2]. Copeland, B.J.. "artificial intelligence". Encyclopedia Britannica, 8 Jan. 2024, https://www.britannica.com/technology/artificial-intelligence. Accessed 9 January 2024.

[3]. Pei, Qin, Luo, Yanan, Chen, Yiyu, Li, Jingyuan, Xie, Dan and Ye, Ting. "Artificial intelligence in clinical applications for lung cancer: diagnosis, treatment and prognosis" Clinical Chemistry and Laboratory Medicine (CCLM), vol. 60, no. 12, 2022, pp. 1974-1983. https://doi.org/10.1515/cclm-2022-0291

[4]. Shigao Huang, Jie Yang, Na Shen, Qingsong Xu, Qi Zhao, Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective, Seminars in Cancer Biology, Volume 89, 2023, Pages 30-37, ISSN 1044-579X, https://doi.org/10.1016/j.semcancer.2023.01.006. (https://www.sciencedirect.com/science/article/pii/S1044579X23000068)

[5]. Xiaomeng Yin, Hu Liao, Hong Yun, Nan Lin, Shen Li, Yu Xiang, Xuelei Ma, Artificial intelligence-based prediction of clinical outcome in immunotherapy and targeted therapy of lung cancer, Seminars in Cancer Biology, Volume 86, Part 2, 2022, Pages 146-159, ISSN, 1044-579X, https://doi.org/10.1016/j.semcancer.2022.08.002. (https://www.sciencedirect.com/science/article/pii/S1044579X22001821)

[6]. Simanta Shekhar Sarmah, Concept of Artificial Intelligence, its Impact and Emerging Trends, Volume: 06 Issue: 11 | Nov 2019, Pages 2164-2168, ISSN: 2395-0056, ISSN: 2395-0072, https://www.irjet.net/archives/V6/i11/IRJET-V6I11253.pdf

[7]. Divya Vemula, Perka Jayasurya, Varthiya Sushmitha, Yethirajula Naveen Kumar, Vasundhra Bhandari, CADD, AI and ML in drug discovery: A comprehensive review, European Journal of Pharmaceutical Sciences, Volume 181, 2023, 106324, ISSN 0928-0987, https://doi.org/10.1016/j.ejps.2022.106324. (https://www.sciencedirect.com/science/article/pii/S0928098722002093)

[8]. Ai X, Guo X, Wang J, Stancu AL, Joslin PMN, Zhang D, Zhu S. Targeted therapies for advanced non-small cell lung cancer. Oncotarget. 2018 Dec 25;9(101):37589-37607. doi: 10.18632/oncotarget.26428. PMID: 30680072; PMCID: PMC6331020.

[9]. Ramesh, P.; Veerappapillai, S. Designing Novel Compounds for the Treatment and Management of RET-Positive Non-Small Cell Lung Cancer—Fragment Based Drug Design Strategy. *Molecules* **2022**, *27*, 1590. https://doi.org/10.3390/molecules27051590

[10]. Schreuder A, Scholten ET, van Ginneken B, Jacobs C. Artificial intelligence for detection and characterization of pulmonary nodules in lung cancer CT screening: ready for practice? Transl Lung Cancer Res. 2021 May;10(5):2378-2388. doi: 10.21037/tlcr-2020-lcs-06. PMID: 34164285; PMCID: PMC8182724.

[11]. Joy Mathew C, David AM, Joy Mathew CM. Artificial Intelligence and its future potential in lung cancer screening. EXCLI J. 2020 Dec 11;19:1552-1562. doi: 10.17179/excli2020-3095. PMID: 33408594; PMCID: PMC7783473.

[12]. Opeyemi Iwaloye, Paul Olamide Ottu, Femi Olawale, Olorunfemi Oyewole Babalola, Olusola Olalekan Elekofehinti, Babatomiwa Kikiowo, Abayomi Emmanuel Adegboyega, Henry Nnaemeka Ogbonna, Covenant Femi Adeboboye, Ibukun Mary Folorunso, Aderonke Elizabeth Fakayode, Moses Orimoloye Akinjiyan, Sunday Amos Onikanni, Sergey Shityakov, Computer-aided drug design in anti-cancer drug discovery: What have we learnt and what is the way forward?, Informatics in Medicine Unlocked, Volume 41, 2023, 101332, ISSN 2352-9148, https://doi.org/10.1016/j.imu.2023.101332. (https://www.sciencedirect.com/science/article/pii/S2352914823001788)

[13]. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Molecules. 2015 Jul 22;20(7):13384-421. doi: 10.3390/molecules200713384. PMID: 26205061; PMCID: PMC6332083.

[14]. K. B. Umare Thool, P. A. Wankhede, V. R. Yella, S. Tamijeselvan, D. Suganthi and R. Rastogi, "Artificial Intelligence in Medical Imaging Data Analytics using CT Images," 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2023, pp. 1619-1625, doi: 10.1109/ICESC57686.2023.10192998.
[15]. https://caseguard.com/articles/the-five-basic-components-of-ai-new-software-development/
[16]. https://www.mygreatlearning.com/blog/what-is-artificial-intelligence/
[17]. https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible
[18]. https://www.analytixlabs.co.in/blog/components-of-artificial-intelligence/