M.Sc. Biotechnology

Harsh Aahra

2023

Machine Learning Analysis Unveil *Withania somnifera* Phytochemicals' Inhibition of GBM through PDK-1 Targeting.

A DISSERTATION

SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

MASTER OF SCIENCE

IN

BIOTECHNOLOGY

SUBMITTED BY

Harsh Aahra
2K21/MSCBIO/16

Under the supervision of

DR. ASMITA DAS

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi college of Engineering)

Bawana Road, Delhi-110042

MAY,2023

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi college of Engineering)

Bawana Road, Delhi-110042

DECLARATION

I, Harsh Aahra, 2K21/MSCBIO/16 of MSc. Biotechnology, hereby declare that the project Dissertation titled "**In Silico and Machine Learning Analysis Unveil Withania somnifera Phytochemicals' Inhibition of GBM through PDK-1 Targeting**" which is submitted by me to the Department of Biotechnology, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Science Biotechnology and submitted to the Department of Biotechnology, Delhi Technological University, Delhi, is an authentic record of my own carried out during the period from January -May 2023 under the supervision of Dr. Asmita Das.

The matter presented in this report has not been submitted by me for the award for any other degree of this or any other institute/University. The work has been accepted in SCI/SCI expanded / SSCI/Scopus Indexed Journal OR peer reviewed Scopus Index Conference with the following details :

Title of the Paper : Diospyros kaki's phytochemical mediated inhibition of GBM by targeting PDK-1: An in-silico docking and machine learning model

Authors Name : Harsh Aahra, Nida E Falak, Asmita Das

Name of Conference : Second IEEE International Conference on "Smart Technologies and System for Next Generation Computing (ICSTSN 2023)

Conference Date and Venue : 21st-22nd April 2023 IFET College of Engineering, Gangarampalaiyam, Tamil Nadu, India

Registration: Done

Status of Paper : Under proceeding

Date of Paper Communication: 20th December 2022

Date of Paper Acceptance: 23 March 2023

Date of Paper Publication:

Date: **Harsh Aahra**

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi college of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "In Silico and Machine Learning Analysis Unveil Withania somnifera Phytochemicals' Inhibition of GBM through PDK-1 Targeting", which is submitted by Harsh Aahra, 2K21/MSCBIO/16, Delhi Technological University Delhi, in partial fulfilment of the requirement for the award of the degree of Masters in Science, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

DR. ASMITA DAS

Supervisor

Date:

ACKNOWLEDGEMEN

I want to express sincere appreciation to everyone who assisted me in completing my thesis. I am eternally grateful for my supervisor, **Dr. Asmita Das**, for their mentorship, generosity, and invaluable insights. This endeavour has benefited greatly from their knowledge and constant support. Additionally, I would like to thank **Professor Pravir Kumar**, **Head of Department** and all the faculty members from **Department of Biotechnology, Delhi Technological Universit**y for their crucial advice and support during my time at the university. My family and friends' steadfast encouragement, sympathy, and enthusiasm are to be greatly appreciated. It has always given me courage to know that they have faith in me.

Place:

Date: **Harsh Aahra**

ABSTRACT

The grade 4 brain tumour glioblastoma multiforme (GBM) is resistant to standard treatments and has a 100% recurrence rate. By blocking the main pathways involved in tumour feeding and development, GBM can be completely eradicated. The AKTmTOR pathway has drawn a lot of attention for GBM therapy because of the higher levels of p-AKT (Phosphorylated Protein Kinase B) seen in recurrent GBM. By being phosphorylated by PDK-1 (3-phosphoinositide-dependent kinase-1), AKT is activated. As a result, PDK-1 activity may be targeted and impaired to stop it from activating AKT. This study uses in-silico analysis to look at different phytochemicals that may be able to target PDK-1. Withania somnifera's phytochemical profile is particularly remarkable in this regard. The high antioxidant capacity of withania somnifera is attributed to the presence of carotenoids, tannins, and phenolic chemicals in the plant. As a result, it is viewed as a viable option for the extraction of anticancer medicinal molecules. The study found four compounds with high binding energies (-11.4, -9.4, -9.8, and -10.4 kcal/mol, respectively) that are comparable to the standard inhibitor 2-(5-[2R]-2-amino-3-phenylpropyloxypyridine-3-yl)8,9-dimethoxybenzo[c][2,7]naphthyridine-4-amine (ID-8I1) compound (-10.1 kcal/mol). Machine learning-based investigations were carried out to evaluate the drug similarity, bioactivity, and bioavailability of the chosen phytochemicals in order to confirm the docking results, indicating their potential as therapeutic agents against GBM. further MACCS descriptors analysis is performed to characterize molecular structure and identify key chemical feature for compound comparison and classification.

Contents

List of Tables

List of Figures

LIST OF ABBREVIATIONS

1. GBM- Glioblastoma Multiforme
2. PDK-1-3-phosphoionsitide-dependent kinase-1
3. GSCs- Glioblastoma stem cells
4. RF- RandomForest
5. RO5- Rule of five
6. NRL- Nuclear receptor ligand
7. ICM- Ion Channel Modulator
8. EI- Enzyme Inhibitor
9. KI- Kinase Inhibitor
10. BBB- Blood Brain Barrier
11. TMZ- Temozolomide
12. MS- Mesenchymal
13. CL- Classical
14. PN- Proneural
15. NE-Neural
16. AC - Astrocyte-like
17. OPC - Oligodendrocytic precursor cell-like
18. NPC- Neuronal progenitor cell-like
19. TCGA- The Cancer Genome Atlas
20. Rb- Retinoblastoma
21. RTK- Receptor Tyrosine Kinases
22. PI3K- Phosphoinositide-3-kinase
23. PIP2- Phosphatidylinositol 4,5-bisphosphate
24. PIP3 - Phosphatidylinositol (3,4,5)-trisphosphate
25. mTORC1- mammalian target of rapamycin complex 1
26. HTS- High Throughput Screening
27. ML- Machine Learning
28. MD- Molecular Dynamics
29. AI- Artificial Intelligence

CHAPTER 1:

INTRODUCTION

Glioblastoma owes its high chances of recurrence and resistance to therapy to a fraction of GBM cells that become highly active and begin to display characteristics resembling those of multipotent stem cells. These cells are referred to as Glioma Stem Cells (GSCs) [1] [2] Several pathways are involved in imparting tumorigenic potential to GSCs [3]. GBM occurrence and resistant to classical drugs like TMZ has been linked to genetic mutated or genetically altered (overexpressed) EGFR [4]or increased phosphorylated AKT levels [2]

Studies have shown that GBM tumour cells have much higher levels of AKT protein phosphorylation, making the P13/AKT/mTor pathway a possible target for inhibition. After receiving the signal from the RTKs, PI3K changes PIP2 into PIP3, which is then changed into AKT, the pathway's most important participant. It cannot serve as a pharmacological target site since it is quite mutagenic. AKT is phosphorylated by PDK-1, a regulatory enzyme that supplies the phosphate group. The mTORC1 pathway is then activated by this phosphorylated AKT, which results in tumorigenic characteristics.[5]

AKT is phosphorylated by a variety of proteins, such as PDK-1, PDK-2, or the TORC2 complex of the mTOR. Although PDK-1 phosphorylation can activate AKT on its own, PDK-2 or TORC2 can help with enhanced phosphorylation[2] *Withania somnifera*-derived phytochemicals will be used to target PDK-1 in this study since PDK-1 promotes oncogenesis and tumour maintenance through AKT [6]. Due to their widespread availability, straightforward extraction techniques, and low toxicity, phytochemicals are currently largely assumed to be anticancer medicines[7]. Withania somnifera is a species of Solanaceae plant that is not only tasty but also has several health advantages. They have the potential to be used as therapeutic agents for cancer cells based on their phytochemical composition and quantity of polyphenols and antioxidants. [8].

The phytochemicals with binding affinities similar to 8I1 were chosen using this compound as a reference. The in-silico screening process was carried out in a number of phases, starting with molecular docking to foresee interactions between phytochemicals and PDK-1 and continuing with machine learning to evaluate and confirm the docking results. The chosen phytochemicals' drug-likeness was assessed using a variety

criteria, including the Lipinski's rule of five, bioavailability radars, and bioactivity scores.

Furthering these efforts, we use MACCS descriptor analysis in combination with machine learning methods. We have effectively deciphered the common chemical functional groups and structural characteristics shared by active medications using the power of MACCS descriptors, enabling us to make educated judgements about compound modification and library creation.

The information gleaned through MACCS descriptor analysis can be used to improve phytochemical compounds' pharmacological characteristics or create chemical libraries. Additionally, by facilitating comparison and similarity searches inside medication databases, this information makes it possible to mine data effectively and find prospective lead compounds. By including MACCS descriptor analysis in drug discovery procedures, researchers are better equipped to decide how to modify compounds and create libraries, which speeds up the creation of new and effective therapeutic agents.

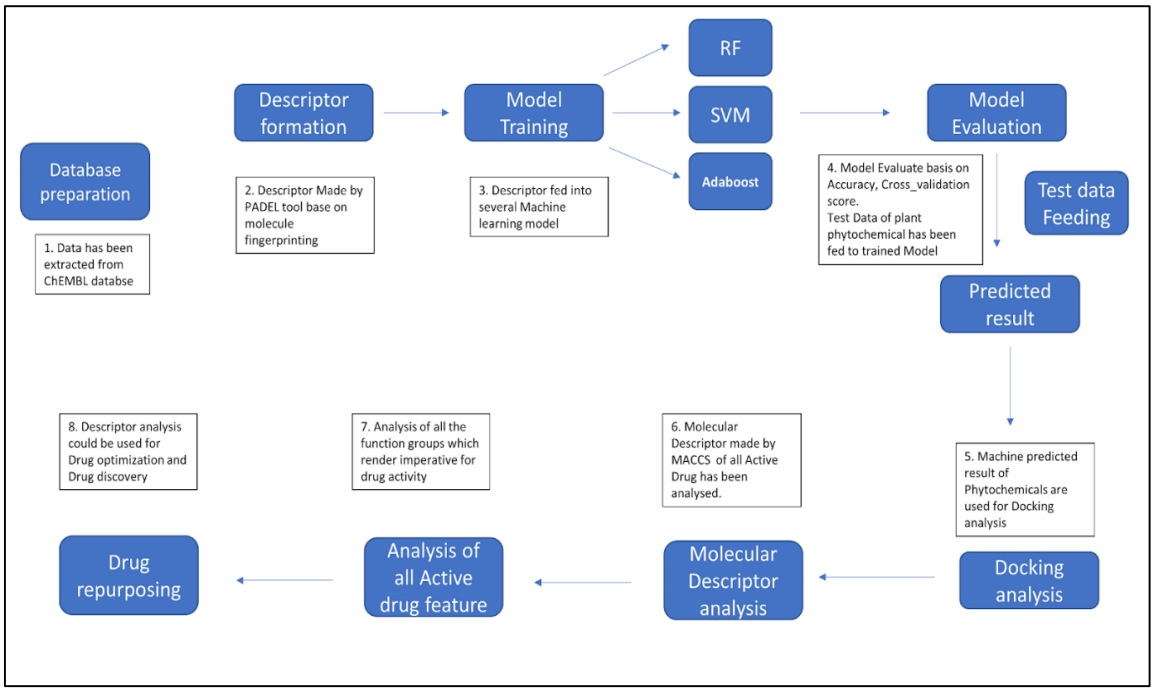


Figure 1: General Workflow of Methodology.

CHAPTER 2:

LITERATURE REVIEW

2.1 General

Cancer is becoming more common and is killing more people. Conventional methods of treatment, such as chemotherapy and radiation therapy, can sometimes be slightly helpful but typically lack specificity and can have major negative consequences. Patients' present treatment options are insufficient due to the complexity and variety of cancers, necessitating targeted therapy or personalised medication [13]. The direct goal of targeted therapy is to impede the biochemical pathways that support the sustenance and growth of tumours. In contrast to conventional methods that may influence both healthy and sick cells, targeted targeting of specific cells minimises the harm to healthy tissues. The molecular and genomic abnormalities that lead to the formation of cancer may now be more precisely identified because to advancements in genome sequencing technology. By using tumour profiling, oncologists can gain a better understanding of the distinct mutations that might cause cancer in various individuals and can consequently administer targeted medicines[13], [14]. To assist patients with treatment choices, personalized medicine incorporates biomarker use such as genetic mutations or protein level expression analysis that can be used to anticipate the reaction of a patient to a therapy or determine the patients who are expected to benefit from a specific drug.

2.2 Glioblastoma

Grade IV tumours like glioblastoma multiforme (GBM) are notorious for having high recurrence rates and being resistant to conventional therapies. A portion of GBM cells become extremely active and start to exhibit traits mimicking those of multipotent stem cells, which contributes to glioblastoma's high likelihood of recurrence and resistance to therapy. Glioma Stem Cells (GSCs) are the name for these cells[1]. Studies show that the transcription factors SALL2, SOX2, OLIG2, and POU3F2 as well as markers like Nestin and CD133, which contribute to the neoplastic activity of GSCs, are abundant in these cells. Nestin, an intermediate filament protein connected to neural stem cells, is a unique feature of neural progenitor cells. The properties of stem cells and the formation of tumours have been connected to the cell surface glycoprotein CD133, also known as Prominin-1 [15] [16].

The aggressive character of GBM, the presence of self-renewing glioma stem cells, the blood brain barrier, the low concentration of medications that may penetrate the BBB, the poor prognosis, and the 100% recurrence rate are some of the key therapy barriers. Based on the bulk tumour transcription profile, expression profiling study has identified four primary categories: Mesenchymal (MS), classical (CL), proneural (PN), and neural (NE) [17]. While OPC (Oligodendrocytic precursor cell-like) and NPC (Neuronal progenitor cell-like) cellular states are prominent in PN and NE, CL and MES tumour types are rich in AC (Astrocyte-like) and MES-like states [18]

Understanding the molecular categories and cellular topologies present in GBM is essential for the development of targeted treatments that may effectively eliminate GSCs and overcome drug resistance. The complexity of GBM, which includes the presence of GSCs and other molecular subtypes, makes it extremely challenging to develop effective treatment regimens. Research is now focused on identifying GSC-specific vulnerabilities and developing medications that can precisely target and eliminate these cells in order to increase survival for GBM patients[16].

2.3 Mutations involved

Table I: Types of Glioblastomas and the mutations involved and their rate of occurrence.

Tumour type	Characteristics	Mutations involved	Occurrence	References
Primary Tumours (90%)	❖ Extremely Aggressive tumours. ❖ Tend to affect the elderly with greater frequency. ❖ Poor prognosis as they develop without any prior symptoms. ❖ Found in frontal and temporal lobes.	P13KCA Loss of RB1 gene IDH1/2 mutation PDGFR amplification GLI1 TERT promoter NF1 deletion/mutation MDM2 PTEN mutation/deletion TP53 mutation CDK2A/B deletion MGMT promoter methylation EGFR amplification LOH at 10q	1% 2% 5% 7% 5-22% 10% 11% 7-12% 24-30% 28-31% 31% 36% 22-40% 65%	[19] [20][21], [22] [20]
Secondary Tumours (<10%)	❖ Less prevalent. ❖ Arise from LGGs (Low Grade Glioblastoma) or AAs (Anaplastic Astrocytoma). ❖ Better prognosis. ❖ Most frequently found in the frontal lobe.	EGFR amplification PDGFR amplification 1p/19q codeletion LOH 19q IDH1/2 mutation TP53 mutation MGMT promoter methylation LOH at 22q	5-7% 7% 15-20% 40-50% 45-50% 65% 75% 70-80%	[20][19] [23] [24]

2.4 AKT/mTOR pathway

According to TCGA, RTK/P13K, p53, and Rb pathways are the biological pathways that are most often affected in glioblastoma [25]. The WNT/-catenin route, however, is also one of the important pathways in GBM, according to certain research, highlighting the strong relationship between the P13K/AKT/mTOR and WNT pathways[26].

Two important factors influencing the aetiology of GBM include lost PTEN function and increased RTKs, such as EGFR. Both of these characteristics apply to the AKT/mTOR pathway, which is negatively regulated by PTEN, which in turn influences PIP3 levels and the pathway as a whole. Since the route cannot be monitored when PTEN function is lost, it gets elevated. Although studies have shown that individuals with GBM have an increased number of receptors (RTKs), which also results in an activated AKT/mTOR pathway[22].

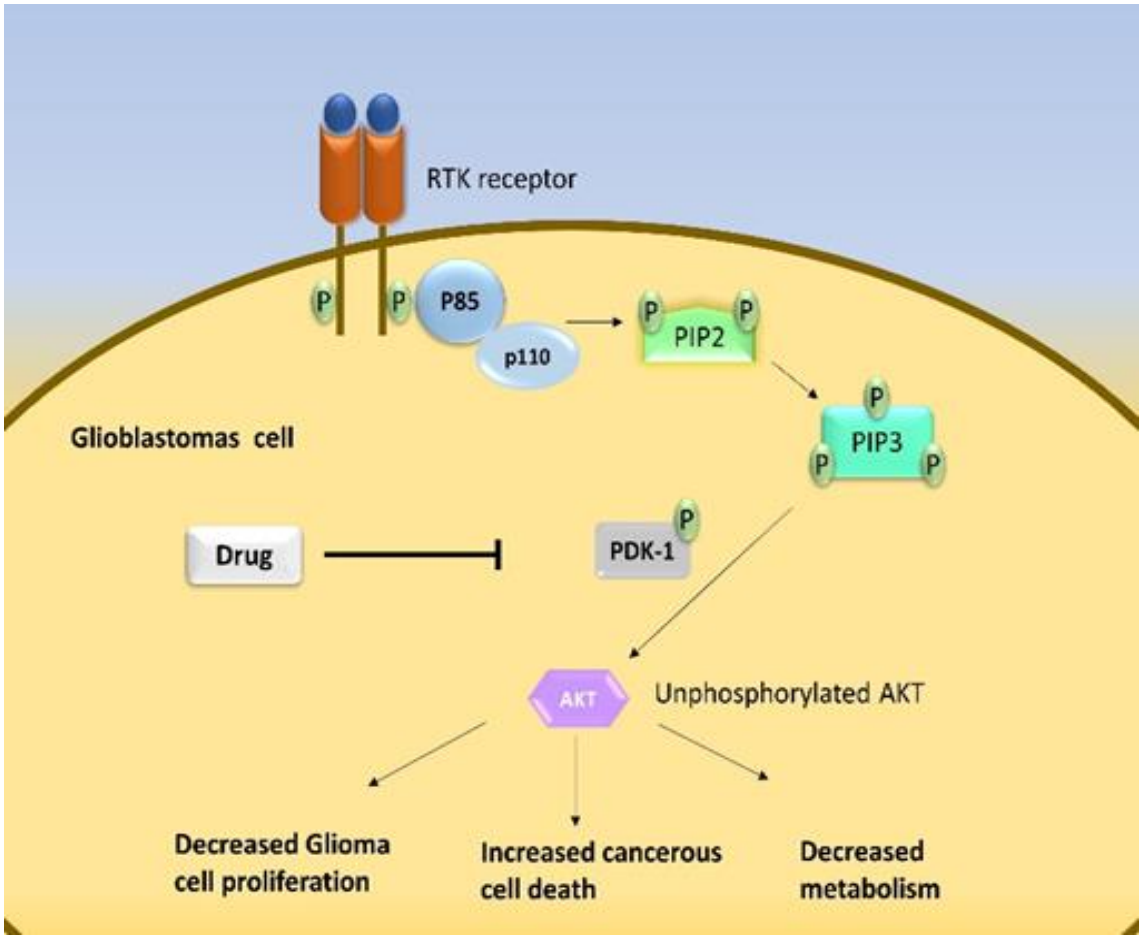


Figure 2: Depicts the phytochemical mediated inhibition of PDK-1 that affects the P13/AKT/mTor pathway in GBM cells resulting in decreased tumorigenic characteristics and eventually death of GBM tumour cells.

Thus, the AKT/mTOR pathway plays a crucial role in giving cells tumor-promoting qualities such as enhanced glioma cell proliferation, increased cellular metabolism, cellular susceptibility, and resistance to conventional therapy (TMZ). Studies have shown that GBM tumour cells have much higher levels of AKT protein phosphorylation, making the P13/AKT/mTor pathway a possible target for inhibition. After receiving the signal from the highly amplified RTKs in GBM, PI3K (Phosphoinositide-3-kinase) changes PIP2 (Phosphatidylinositol 4,5-bisphosphate) into PIP3 (Phosphatidylinositol (3,4,5)-trisphosphate), which is then converted to AKT, the pathway's most important player. It cannot be employed as a direct drug target site since it is extremely mutagenic. AKT is phosphorylated by PDK-1, a regulatory enzyme that supplies the phosphate group. This phosphorylated AKT in turn activates the mTORC1 (mammalian target of rapamycin complex 1) pathway and induces tumorigenic properties [27]

2.5 PDK-1

AKT is phosphorylated by the serine-threonine kinase PDK-1, which functions as a regulator. According to studies, PDK-1 and NFkB help GBM cells fight and escape the effects of the medication TMZ. The Warburg effect, which occurs whether or not there is enough oxygen available, causes GBM cells to preferentially create lactate rather than oxidative phosphorylation. Since pyruvate must be converted to lactate for PDK-1 to function, the oxidation of glucose is checked at this step[28]. the reversal of the Warburg effect and reduced phosphorylation of the AKT protein, which in turn diminishes the tumorigenic characteristics and tumour resistance to the conventional TMZ. Therefore, PDK-1 of the AKT/mTOR pathway has been chosen as the prospective therapeutic target in this investigation.

Researchers from all around the world are now focusing more on employing natural products or phytochemicals to treat cancer due to a rise in the amount of side effects and hazardous responses caused by synthetic drugs. In order to treat patients who have no other accessible options for treatment, there is a high demand for researching previously untapped natural resources in order to produce treatments against particular diseases. Phytochemicals owe their potential to be used as drugs against cancer or as adjuvant therapy to their non-toxicity, their easy availability and extraction, their efficacy, safety as well as mechanisms of action.

2.6 Withania somnifera

Withenia somnifera also famously known as Indian ginseng has great potential for treatment of various disease and promoting health benefits. It contains a number of different potent phytochemicals enable it to be utilize in significant biological contexts. Alkaloids like anahygrine, anaferine, isopelletierine and cuseohygrine are among the several physiologically active chemical components found in Withania somnifera (WS). it also includes steroidal lactones such withanolides and withaferins as well as saponins. Additionally, ashwagandha includes acylsterylglucosides and sitoindosides, two anti-stress compounds. These chemicals have significant ability to reduce the negative impact of stress. The pharmacokinetics studies of various phytochemicals show rapid absorption and plasma half-life of 1.36h.

One of its aspects is inhibiting cancer growth and proliferation in various tumor like colon, breast, lung cancer cell and brain cancer. There is various mechanism of inhibition of cancer, which involves anti-inflammatory, immunomodulatory. Antiangiogenic and anti-mitogenic effect. Phytochemical of *Withenia somnifera* seen to be target various protein molecules such as STAT3, CDK-1, MAPK and COX-2, P53 inactivation. Most of these proteins have been identified to be upregulated in brain cancers.

2.7 Molecular Docking

It is feasible to study a molecule's pose—or structure and location—during its interaction with a target's binding site by looking at a procedure known as molecular docking. This technology has transformed drug design and optimisation by enabling simulated testing of compounds, which lowers costs and speeds up the discovery process. The hit-to-lead optimisation process's primary method for screening compounds, High Throughput Screening (HTS), is based on molecular docking theory. A scoring function that assesses the degree of contact between the molecule and the target and structural and orientation-based search algorithms that determine where the molecule should be positioned within the binding site are two essential components employed in the docking process. These elements work together to make it easier to find and improve new medication candidates. [31]

For molecular docking, the software programmes Molegro Virtual Docker, FlexX, DockThor, GOLD, AutoDock, and AutoDock Vina are often used. AutoDock Vina and AutoDock Perl were employed in this inquiry due of their ability to do rapid and accurate analysis. Several databases, including as PubChem, ZINC, and PDB, include the 3D structures for the ligands. For this experiment, the PDK-1 protein structure was obtained from the PDB database, whereas phytochemical structures were obtained from the IMMPAT database [31], [32].

The potential for medical research may be greatly expanded by integrating molecular docking with a variety of computational biology fields, including as machine learning, artificial intelligence, molecular dynamics, pharmacokinetic analysis, and bioavailability analysis. By identifying potential interactions between molecular targets and ligands, molecular docking makes it easier to identify new targets for existing ligands, predict potential negative effects of medications, and investigate the repositioning and repurposing of FDA-approved drugs. The procedure is more successful overall because to this comprehensive approach, and it also opens up new possibilities for field advances.

Table II: Integration of Molecular docking with other computational branches resulting in the improved efficiency of the whole process.

	Integrated fields	Pre-docking Screening	Post-docking Screening	References
Molecular Docking	AI and Statistical Approaches	Retrieval of protein structures for screening Optimizing the score	Pose improvement	[31], [32]
	Molecular Dynamics (MD)	Selection of conformations	Phytochemical-target interaction evaluation. Orientation and conformation refinement	[33]
	Ligand-based methods	Retrieval of protein structures for screening	Orientation and conformation selection and scoring	[31], [32]
	Binding Free Energy methods		Orientation and conformation refinement	[32]

2.8 Standard/Inhibitor used

To serve as a reference for the molecular docking scoring analysis, a specific inhibitor is retrieved from the PDB database. The comparative analysis of binding affinities of the target-inhibitor complex and the target-ligand complex provides an insight about the extent to which a particular ligand is interacting with the target protein. This study has selected the compound 8,9-dimethoxy-5-(2-aminoalkoxy-pyridin-3-yl)-benzo[c][2,7]naphthyridin-4-ylamine as the standard inhibitor. [34]

2.9 Machine Learning

Python is a popular programming language in the machine learning industry because of its adaptability, sizeable library, and user-friendliness. Using current data from the ChEMBL database can be quite beneficial when it comes to evaluating the outcomes of machine learning models in the context of ligand-drug interactions. ChEMBL is a comprehensive tool that gives users access to a huge library of bioactive compounds, their characteristics, and related biological functions. This database offers data on numerous medications and how they interact with various proteins. The information includes a variety of characteristics, such as ChEMBL IDs, molecular weights, and SMILES sequences. (a compact representation of molecular structure), and IC50 values (a measure of drug potency). The evaluation of in-silico docking findings can be improved by using the ChEMBL data to train machine learning models. With the aid of this dataset, researchers can train their models to determine whether or not the interactions between ligands and target proteins are accurately reflected in their docking results. [35]

The procedure normally entails pulling pertinent information from the ChEMBL data, including molecular weights, SMILES sequences, and Molecule ChEMBL id. The size and mass of the ligands are revealed by their molecular weights, which can reveal whether or not they have the ability to interact with proteins. SMILES sequences, in contrast, encode the ligands' chemical structures, enabling a more thorough examination of their possible binding abilities. The collected features may then be paired with the associated IC50 values, which function as the ground truth or labels when training a machine learning model. The IC50 values show the amount of a medicine needed to completely block a certain biological process. This knowledge can help the model learn to link certain characteristics to desirable binding affinities throughout the training phase. It is crucial to

divide the ChEMBL dataset into training and testing sets in order to guarantee the machine learning model's dependability and generalizability. The model is trained using the training set, and its performance on test data is assessed using the testing set. This procedure aids in evaluating the model's propensity to generalise and deliver precise forecasts for fresh docking outcomes. [36]

The machine learning model may be used to predict the binding affinity or IC50 values for brand-new ligand-protein interactions after it has been trained. The accuracy of the docking results may be evaluated by comparing these predictions to experimental data or known IC50 values from the ChEMBL database. This validation stage offers insightful information on the dependability and efficiency of the in-silico docking method.

2.10 MACCS fingerprints

In the field of cheminformatics, one popular technique for producing molecular descriptors or fingerprints is the Molecular ACCCESS System (MACCS). The presence or absence of preset substructures or structural properties in a molecule is represented by MACCS keys, which are binary fingerprints. The 166 keys or bits that make up a MACCS fingerprint each stand for a distinct chemical characteristic. Aromatic rings, functional groups, atom kinds, and other structural traits are among the qualities represented in MACCS keys. If the matching characteristic is present in the molecule, each bit is given a value of 1; otherwise, it is given a value of 0 (zero). [37]

Chemical similarity searches, virtual screening, and drug development are just a few of the uses for MACCS fingerprints. Researchers can find related compounds or sift huge chemical datasets to prioritise molecules with desirable structural properties by comparing the MACCS fingerprints of various molecules. A chemical structure is examined and transformed into a binary fingerprint using preset procedures for each MACCS key to produce MACCS fingerprints. With the use of similarity metrics like the Tanimoto coefficient or Euclidean distance, the generated fingerprint effectively compares to other fingerprints by providing a condensed representation of the structural characteristics of the molecule. Overall, MACCS descriptors offer a method for comparing and capturing molecular structure data in a condensed binary format, which makes computer analysis and prediction jobs in cheminformatics and drug development easier. [38]

Purpose of MACCS analysis

Drug research and development are greatly aided by the examination of effective medications using MACCS fingerprints because it sheds light on the essential characteristics and structural characteristics of effective therapeutic agents. In order to repurpose currently available pharmaceuticals or create new ones with increased efficiency, researchers can find common functional groups, rings, or metals that are commonly present by analysing the fingerprints of known effective drugs.

The capacity of MACCS fingerprints to represent chemical structures in a small binary representation is one of its main features. Large chemical datasets may be stored, compared, and retrieved effectively using this format. Researchers can find recurrent substructures or traits that are connected to therapeutic efficacy by comparing the MACCS fingerprints of known effective medications. These characteristics, which might have a significant impact on the drug's mode of action or target interaction, may include certain functional groups, aromatic rings, or even the presence of certain metals.

Researchers learn more about the structural and functional traits that contribute to a drug's efficacy through MACCS analysis for instance, if a certain functional group frequently appears in effective medications, this may indicate that this group is crucial for target binding or particular biological interactions. Using this knowledge, medicinal chemists can concentrate their efforts on adding or changing these essential functional groups to current medications or when developing new drug candidates. [39]

The existence of particular metal ions may be inferred from MACCS fingerprints in addition to functional groups and rings. Metal-containing medications have demonstrated substantial therapeutic benefit, such as platinum-based chemotherapeutic treatments or metalloenzyme inhibitors. Researchers can use the predominance of particular metal characteristics in effective pharmaceuticals to guide the creation of novel metal-based treatments or aid in the optimisation of currently available metal-containing drugs by analysing MACCS fingerprints.[39]

Overall, an organised and effective method for identifying the structural and functional characteristics of effective medications is the MACCS analysis. Researchers can repurpose already available medications, create new ones with better attributes, or concentrate on particular functional groups, rings, or metals that are crucial for

pharmacological action by discovering the recurrent characteristics linked with their efficacy. By rationally designing and optimising medication candidates with the use of this information, numerous diseases can eventually be treated more successfully.

CHAPTER 3: MATERIALS

3.1 IMPPAT database

Indian Medicinal Plants, Phytochemistry and Therapeutics is a sizable resource that focuses on the medicinal plants found in India. IMPPAT may be very helpful to researchers, scientists, medical professionals, and anybody else interested in learning about or applying the medicinal properties of Indian plants. A variety of plant species are covered in-depth in the database, including their botanical names, common names, traditional uses, chemical make-up, pharmacological activity, and relevant scientific investigations. In view of the growing demand for natural products and conventional medicine, IMPPAT is essential in promoting the discovery and usage of Indian medicinal plants. It facilitates research on the phytochemical evaluation of plant extracts, the identification of bioactive compounds, and the development of plant-based therapies.[40]

IMPPAT offers a comprehensive library of data on the phytochemical composition of Indian medicinal plants, which helps the in-silico drug discovery process. This resource contains details about the chemical elements present in many plant species. By using computer analysis, researchers may predict the likely biological properties and activities of these substances, such as how they interact with certain proteins or enzymes.

The database also makes it easier to study structure-activity correlations (SAR) using computer modelling. By examining the chemical structures of bioactive compounds and their associated biological activities, researchers may develop prediction models to foresee the biological activities of analogous substances. This allows for the discovery of intriguing molecules and their subsequent production or modification to enhance their properties, boost their efficacy, and minimise their toxicity. [40]

Additionally included in IMPPAT's in silico capabilities are molecular docking studies and virtual chemical libraries. The database might serve as a starting point for the creation of digital collections of chemical compounds obtained from plants. These libraries can be

used for screening against specific therapeutic targets or for virtual docking investigations where computer algorithms predict the binding affinity between a chemical and a target protein. Such simulations aid both the identification of potential drug candidates and the enhancement of their binding interactions.

3.2 PDB

The protein data bank is a sizable global store of structural information for big molecules like protein and nucleic acid. It was formed in 1971, but there wasn't enough data accessible until 1980. As technology advanced via the use of crystallography, nuclear magnetic resonance (NMR), cryoelectron microscopy, and theoretical modelling, the number of deposited structures began to rise. PDB is crucial for structural genomics, structural biology, and bioinformatics. Numerous programmes, such CATH and SCOP, employ protein structures that are stored in the PDB. PDB holds more than 180000 macromolecule structures, including many 3D structures of nucleic acids and proteins.[41]

It provides an extensive and varied variety of protein structure and enables researchers to examine an interaction, function, and their folding, making it a useful resource for students, researchers, and educators throughout the globe. Structure information has been determined and obtained using a variety of methods, including crystallography, magnetic resonance spectroscopy, and electron microscopy. To verify the accuracy and dependability of the data, many stringent validation processes have been used

PDB includes ligand and small molecules that are essential to biological activity and process in addition to protein and nucleic data. PDB is particularly useful for researching protein-protein interactions, complicated molecular interactions, finding new drugs, creating enzymes, and basic biological mechanism investigations. The PDB also offers a number of tools and resources to help with data analysis and visualisation. PDB offers a dynamic molecular viewer, search features, and sophisticated query options. These technologies aid scientists and researchers in understanding the functional characteristics of nucleic acid and protein molecules by allowing them to compare molecular structures.[41]

3.3 Autodock vina

In order to predict the score and mode of binding, protein-ligand docking analysis is performed to ascertain a ligand's affinity to a protein molecule. Since its first release in 1970, Autodock Vina has sparked active research efforts, aided in the creation of new medications, and enhanced ones already on the market.

Docking programmes are a methodology based on a scoring function and an exploration strategy for exploring and sampling both the positional and structural space. An already-programmed system's free energy is assessed using the scoring function.AutoDock Vina, a well-known piece of software, is crucial for computational drug development and virtual screening. [42]

One of AutoDock Vina's primary strengths is how quickly and precisely it searches the vast conformational and positional space. It employs a hybrid search algorithm, which combines techniques for both global and local optimisation, allowing for a detailed analysis of ligand binding poses. By autonomously sampling a range of ligand orientations and conformations, AutoDock Vina considerably enhances the likelihood of finding energetically favourable binding configurations.[33], [43]

When creating Vina, researchers looked into a range of stochastic global optimisation techniques. These methods included genetic algorithms, particle swarm optimisation, and simulated annealing, among others. Additional local optimising processes and specific optimisation techniques were applied to hasten the process. Vina chose the Iterated Local Search global optimizer as its tactic after doing extensive study. [33], [42]

The algorithm employs an empirically computed set of parameters and energy terms to create a score that represents how effectively the ligand and receptor bind. The scoring function used to determine the binding energy takes into account intermolecular forces such hydrogen bonds, van der Waals contacts, and electrostatics. By considering these factors, AutoDock Vina can effectively rank different ligand poses in accordance with their predicted binding affinities.

Additionally, AutoDock Vina offers a user-friendly interface and a wide variety of movable options that let researchers tailor the docking process to their own needs. Due to its ability to include many molecular forms, it is compatible with a range of ligand and

receptor combinations. With the use of the software's visualisation tools, which are also supplied for studying and interpreting the docking data, researchers may identify potential binding sites and appreciate how ligands and proteins interact[44].

3.4 PLIP

Protein-Ligand Interaction Profiler (PLIP), a powerful computer technique, is used to study and analyse how proteins interact with their small molecule ligands. Due to the evolution of structural biology and the increasing accessibility of protein-ligand complex structures, it is crucial for drug development, molecular biology, and bioinformatics study to comprehend these interactions. PLIP, which is both comprehensive and user-friendly, assists in characterising the intricate interactions between proteins and ligands. It examines the affinities, sites, and behaviours of ligands as they bind to proteins using a range of methods and methodologies. The three-dimensional nature of these interactions is crucial information gained via the use of a variety of structural analysis tools, and PLIP helps researchers clarify their functional and mechanistic consequences.[45]–[47]

One of PLIP's important features is its ability to detect and classify ligand-binding regions on protein structures automatically. It looks at the protein's surface, finds potential binding sites, and then uses sophisticated algorithms to group them into different ligand classes. The capacity to swiftly identify potential binding sites and prioritise them for further analysis expedites the drug development process.

PLIP also employs molecular docking techniques to predict the binding poses of ligands inside protein structures. By computationally determining the ideal spatial arrangement between the protein and ligand, these docking simulations provide crucial insights into the binding affinities and likely interaction pathways. This information makes it easier to develop novel ligands rationally and to improve existing treatment possibilities.[46]

Additionally, PLIP offers a robust visualisation feature set that makes it easy for researchers to study protein-ligand interactions. By highlighting the essential residues involved in ligand binding and providing a complete grasp of the binding modes, it provides dynamic 3D protein structure visualisations. These visual representations facilitate the dissemination of research findings to a larger scientific community and make it simpler to comprehend complex linkages.

Any structure from the RCSB PDB service may be processed using a four-letter PDB Id, a free text search in the ligand-protein complex, or by loading a PDB file into PLIP. In-silico docking output files can also be supplied to PLIP for analysis of covalent and non-covalent interactions.

The PLIP output generally analyses the ligand-protein complex. In addition to a table describing the protein amino acids involved in covalent or hydrogen interactions, PLIP provides a 2D and 3D interaction graphic for each ligand-protein interaction. The output file from PLIP is also available in PNG and PyMOL formats. You may get the details of the interaction pattern by clicking the diagram's high-level view. [46]

3.5 BioVia Discovery studio

Exploring and fusing art and science is made simpler with a ground-breaking application called Visual Discovery Studio. It provides a collaborative setting for designers, scientists, and artists to collaborate on complex ideas, facts, and thoughts. The company creates visually appealing and factually accurate visualisations using a variety of methods and technologies, including digital design tools and traditional art supplies.[48]

Visual discovery studio is an effective tool for visualising protein-ligand interactions. In preparation for docking analysis, the protein and ligand are visualised. In order to make proteins, a poler hydrogen group must be added, water must be removed from protein molecules, and extra peptide or ligand chains must be cut. We can distinguish between the many types of bonds that the ligand and protein molecules create using the 2D depiction of interaction that Visual Discovery Studio provides.

3.6 ChEMBL Database

ChEMBL is a helpful and widely used database in the field of medicinal chemistry and drug development. It provides a comprehensive collection of information on bioactive substances, their targets, and the biological processes associated with them. This resource has substantially sped up the process of developing new medications by making it possible to conduct research on chemical compounds and their interactions with biological systems.

ChEMBL is a substantial, open-access drug database that seeks to gather data from the healthcare and pharmaceutical sectors as well as from the process of studying and creating

medications. A large number of medications and ligands have been tested on a variety of proteins and biomolecules using ChEMBL. It keeps track of small molecule and medication information as well as data on their biological activities from several medical chemistry publications. To give researchers access to comprehensive information, bioactivity data are shared with other databases including BindingDB and PubChem Bioassay[49]

Research papers from a range of publications, such as Journal of Medicinal Chemistry, Bioorganic Medicinal Chemistry Letters, and Journal of Natural Products, are mined for vital activity data. The chosen journals have been carefully chosen to ensure the efficient use of resources while acquiring a substantial amount of reliable data, even though they do not cover every scenario. Each article's database abstracts provide details about the tested chemicals, the experiments that were conducted, and any pertinent target data.

ChEMBL provides researchers with access to a multitude of data, such as details on pharmacological profiles, binding affinities, and compound structures. The database includes data from a variety of sources, including public databases, patents, and scientific publications. With its extensive coverage and continuous updates, ChEMBL offers researchers a robust platform for identifying potential drug targets, investigating structure-activity relationships, and designing new compounds. [49]

One of ChEMBL's main benefits is its user-friendly interface, which makes it simple for researchers to search for and collect data. Users have the ability to conduct advanced searches, filter results using predetermined criteria, and obtain thorough annotations for certain substances and targets. The database's capabilities for data visualisation and analysis allow researchers to get insights from the vast amount of information available. ChEMBL has significantly aided in the advancement of drug development and research. By merging information from many sources, it has grown to be a valuable tool for researchers all around the world. The database promotes transparency and collaboration by making its contents freely accessible, allowing researchers to use the most recent information and build upon prior work.[50]

Data access form ChEMBL

Data retrieval is made easier by the ChEMBL interface's simplicity. Users can input a keyword, protein name, ChEMBL target identifier, or UniPort accession of an interesting target for which a ligand needs to be found in the interface's search tool.

Once users have obtained a target or a number of targets of interest from the ChEMBL database, they may rapidly access the supporting bioactivity data using a drop-down menu. Using this user-friendly capability, customers may view all the data that is accessible or create filters to choose only particular activity types. Users can choose, for example, to focus on specific ADMET endpoints or only include IC50 and Ki data below a specific concentration threshold. [50]

The following bioactivity table, which also contains information on the specific salt form used in the experiment, provides a detailed description of each studied medication. Additionally, it comprises details about the test, a description of the target (including the organism), as well as its description, units, and the type of observed activity. It should be noted that the table includes a link that points directly to the article from which the data were derived, ensuring accessibility and transparency to the original source. Researchers may immediately export the data from this view as a text file or spreadsheet to make additional research and analysis easier. Users may utilise this tool to go further into the data, conduct their own study, and extrapolate significant findings from the discovered bioactivity data.[51], [52]

3.7 Swiss ADME

Chemically synthesising, developing, testing, and optimising a medicine requires access to a number of factors, including biological activity, toxicity, concentration, etc. The chance of clinical-phase ADME (Absorption, Distribution, Metabolism, Excretion)-related failures is greatly reduced by pharmacokinetic evaluation at the start of the discovery phase. Swiss ADME is a well-known online tool for its dependability and robustness as well as for its straightforward result analysis that makes efficient incorporation to drug development via molecular design possible.

The programme offers a wide range of input possibilities, including chemical structure and canonical smiles, analysis of different compounds, the ability to store and share results, and user-friendly interactive graphs like the boiled egg and the bioavailability radar. The cooked egg is useful in assessing BBB penetration and gastrointestinal absorption. The yellow yolk region displays the substances that have a high BBB permeability, whereas the white region, also known as the albumin region, displays the substances that are most likely to be passively absorbed via the gastrointestinal system. On the other hand, the bioavailability radar has a pink region that shows the ideal range of characteristics, including flexibility, saturation, solubility, polarity, molecular weight, and lipophilicity. A compound's radar should fall in the pink area for it to be considered as a good drug candidate. [53]

3.8 Molinspiration

Numerous computational tools are available on the Molinspiration software platform, which is designed to help with molecular analysis and modification. These technologies support virtual screening, data visualisation, and bioactivity prediction. The bioactivity score expectation for therapeutic targets (GPCR, KI, ICM, and NM scores) is the most crucial analysis performed by Molinspiration in drug development and optimisation.

For these studies, Molinspiration accepts input files in the SMILES or SDfile formats. SMILES (Simplified Molecular Input Line Entry System), a string-based syntax, and SDfile (Structure-Data File), a file format often used to store and communicate chemical structures and related data, are used to express the molecular structure of a substance. In order to determine the bioactivity scores for the selected therapeutic targets, Molinspiration applies its computational methodologies to the input file that has been supplied. In order to assess a molecule's potential as a therapeutic candidate, these grading systems provide information on the molecule's potential activity or affinity towards the target. By prioritising compounds for more experimental testing, this capability helps speed up the drug development process.

3.9 Lipinski's rule of five

The permeability and oral bioavailability of a medicine are two very important parameters that help determine a compound's drug similarity or effectiveness. The Lipinki's criteria are based on research that shows that characteristics of orally delivered drugs that are readily absorbed fit within these restrictions: 500 Da in mass, 5 H-bond donors, 10 H-bond acceptors, and a 5 LogP value.

The likelihood of unfavourable pharmacokinetic characteristics such quick metabolism, decreased availability, decreased permeability, and failure to penetrate cell membranes is higher for substances that stray from these limits. In order to screen out and rank compounds according to their propensity to be orally active, the rule of five is a gold standard procedure in drug development and optimisation techniques.[54]

CHAPTER 4: METHODOLOGY

A. Selection and preparation of ligands/Retrieving phytochemicals and preparation of drug library

Forty bioactive compounds from *Withania somnifera* were collected from IMPPAT database in pdb format [40]. Using Open Babel GUI, pdbqt formats of the files were obtained.

B. Protein Preparation

The structure of target protein PDK-1 (3-Phosphoinositide-dependent kinase 1) was retrieved from PDB. Along with the target protein structure, the structure of PDK-1 inhibitor 8I1 [34]was also retrieved to be used as a reference[41].

Using Autodock Vina 1.7.5, Water molecules were removed as they generally do not participate in the binding process. So, in order to simplify the computational calculations and get rid of any potential obstructions and pose distortion in the binding pocket.Polar hydrogens along with Kollman's charges were added as the pdb files lack hydrogens, so in order to attain accurate optimization and calculations, charges and hydrogens were added and pdbqt format of the protein was retrieved. Open Babel GUI was used to convert inhibitor 8i1 from pdb format to pdbqt format.

C. Molecular docking

- **Protein-ligand docking using Auto dock Vina**: Autodock Vina 1.5.7 employed a computational docking strategy by setting the x, y, and z centre dimensions to -35.04, -26.046, and -1.741, respectively, and the x, y, and z sizes to 58, 48, and 74, respectively, docking analysis was performed. Number of modes and energy ranges were set to 10 and 4 respectively. All phytochemicals are docked against the target protein by using Auto dock Vina and Perl. The standard inhibitor 8i1 was also docked against target protein [44].

- **Docking/Interaction analysis**: The downloaded outputs from auto dock were analyzed via PLIP and Discovery studio to identify by which amino acid the ligand binds to the protein. The binding energies of ligands and standard were compared and phytochemicals with high binding energies were selected [46], [48].

D. **Pharmacokinetic and Drug Likeness Screening of selected Phytochemicals:** Only phytochemicals that can get past the initial round of binding energy range screening are subjected to further analysis based on their drug likeness and pharmacokinetics.

- **Lipinski's RO5 analysis:** The rule of five assesses drug likeness, or the likelihood that a molecule will be active when taken orally. The selected phytochemicals were subjected to Lipinski's RO5 to analyze their oral activity [54].
- **In-silico bioavailability analysis:** To comprehend the pharmacokinetics of a drug, it's vital to understand its absorption, distribution, metabolism, and excretion. In order to evaluate the bioavailability radars of the phytochemicals, SwissADME and admetSAR[53] were employed. By providing canonical SMILES of phytochemicals as input, these values can be determined [53].
- **Bioactivity score:** To determine the druggability characteristics of ligands like NRL, PI, and EI, GPCR, ICM and KI, bioactivity score is required. The scores can be predicted by providing canonical SMILES of phytochemicals as input to

  Molinspiration**. (**Molinspiration Cheminformatics free web services, https://www.molinspiration.com, Slovensky Grob, Slovakia)

E. **Machine learning Validation**

1. **Data extraction:** The initial step in machine learning is data extraction. In this case, the data from the ChEMBL database was extracted using the Python tool chembl_webresource_client. The data extraction involves the protein PDK-1, and all ligands that have been properly researched and examined in relation to this protein have been created. There are 1150 ligands in the gathered dataset. After the data extraction is complete, the dataset is filtered using the IC50 values. The IC50 value of a medicine, which displays the dose required to 50% block a certain biological process, is used to assess its potency. In order to identify ligands that significantly interact with the protein PDK-1, the data may be filtered based on IC50 values. By filtering the dataset, one may make sure that the machine learning model focuses on ligands that are more likely to be effective and relevant for the target protein. By focusing on ligands with higher inhibitory potency against the target protein, researchers can prioritise them by limiting the dataset based on IC50 values. [50] To filter the data, several IC50 filtering criteria may be used, depending on the requirements and objectives of the study. Researchers may choose to include ligands with IC50 values below a specified threshold, implying higher potency, or they may focus on a specific range of IC50 values depending on the desired level of protein interaction. It is important to keep in mind that the dataset's quality and suitability for subsequent machine learning operations are guaranteed by the data extraction and filtering processes. By eliminating data specifically related to the target protein and filtering using IC50 values, researchers may create a more condensed and curated dataset that is more suited for training and validating machine learning models.

2. Data Classification: The next step involves categorising the ligands according to their IC50 values once the data from the ChEMBL database has been extracted. In this instance, the distinction between active and inactive drugs is made using a threshold of 100. A ligand is categorised as active if its IC50 value is less than 100; otherwise, it is categorised as inactive. Which drugs are more successful at inhibiting the target protein may be ascertained with the use of this classification procedure. By categorising the ligands as either active or inert based on their IC50 values, researchers can get additional knowledge about the potential usefulness of these drugs. Evaluation of any data bias in

the dataset is also crucial. Any systematic inaccuracy or irregularity in the dataset that might influence the findings or forecasts produced by the machine learning model is referred to as data bias. Data bias needs to be taken care of to guarantee the model's reliability and generalizability. To check for data bias, researchers frequently do a thorough analysis of the dataset. The distribution of active and inactive medications, the distribution of IC50 values, and any potential sources of bias in the data gathering procedure are just a few of the variables they look at.

3. Making descriptors: Based on the ligands' SMILES sequences, the PaDEL software has been used to create molecular descriptors or fingerprints. A popular piece of software for computational drug design and research is PaDEL. Prediction and Evaluation of Drug-likeness and Toxicity Liability is what it stands for. It offers a variety of molecular descriptors to explain various aspects of a molecule's structure and properties. In this case, 308 descriptors were produced by the PaDEL programme using the structural characteristics of the ligands determined from their SMILES sequences. Numerous molecular features, including as size, shape, flexibility, electrostatics, and other structural traits, are numerically represented in these descriptions. They provide useful information that machine learning algorithms may use to identify patterns and connections between molecular structures and their biological activity.. Molecular descriptors are crucial in the discovery and creation of novel medications. They enable researchers to compare and assess the chemical characteristics of various compounds effectively in order to predict a variety of molecular and biological attributes, such as drug-likeness, solubility, toxicity, and activity against target proteins. By incorporating chemical descriptors produced by the PaDEL into the machine learning process, researchers may train algorithms to analyse and detect patterns in the data. These models may then be used to predict the activity or potency of new ligands based on their chemical structures in order to find potential therapeutic possibilities. [8]

4. Machine Training: The dataset was divided in half throughout the machine learning training phase, with 80% of the data used to train the model and 20% used to assess its performance. It is feasible to evaluate how well the model generalises to fresh data thanks to this split. The data have been trained using the RandomForestClassifier model in this specific instance. preferred method of machine learning A collection of decision trees are used by RandomForest to forecast results. It is well known for having

the ability to handle complex connections and produce consistent results. [55]. After the model has been trained with the training data, the accuracy of the model is evaluated using the testing data. Accuracy is a common measure used to gauge a classification model's performance. Out of all the occurrences in the testing dataset, it calculates the percentage of instances that were properly categorised. In this case, the test data performance of the RandomForestClassifier model was determined to be 85% correct. This shows that in 85% of the cases, the model correctly predicted the ligands' activity or inactivity based on their chemical descriptors and other properties[56]. The model's accuracy of 85% demonstrates how well it predicts the ligands' actions. It is important to undertake a more in-depth investigation and consider additional evaluation criteria in order to properly understand the model's performance. Other well-known evaluation criteria for classification problems include precision, recall, and F1 score. Precision is the proportion of precisely predicted positive events among all cases that are expected to be positive. Recall, sometimes referred to as sensitivity or the true positive rate, is the ratio of correctly predicted positive cases to actual positive occurrences.

The harmonic mean of recall and accuracy is the F1 score, which is a balanced assessment of model performance. [Progress with diffuse large B-cell lymphoma] To gauge the model's robustness and generalizability, cross-validation ought to be utilised. The dataset is split up into several subsets for cross-validation, and the model is trained using a variety of pairings of training and testing sets. By making sure the model functions consistently across different data divisions, this reduces the likelihood of model failure. ElasticNet, RandomForestClassifier, NuSVC, BaggingClassifier, HistGradientBoostingClassifier, SVC, RidgeClassifier, DecisionTreeClassifier, KNeighborsClassifier, MLPClassifier, and AdaBoostClassifier are just a few of the models that have been used in the machine learning process to train and evaluate the data. Among these models, the RandomForestClassifier has provided the best cross-validation result.

5. Phytochemical data preparation and validation: The IMPPAT database had 80 phytochemicals from the Withania somnifera plant that were submitted. The chemical structures of these drugs were represented in an Excel spreadsheet using SMILES sequences. Molecular fingerprints or descriptors were developed using the PaDEL tool in order to predict the activity of these phytochemicals and evaluate their potential as inhibitors. This technique generated a total of 308 descriptors for each compound, each

of which described a unique collection of structural and physicochemical properties. The generated descriptors were used to build a machine learning model for activity prediction by feeding it into a RandomForestClassifier. To provide accurate predictions, this classifier combines many decision trees, an effective ensemble learning approach. It has been heavily used in chemoinformatics and drug discovery applications because to its ability to manage complex interactions. The trained RandomForestClassifier model was then used to predict the activity of the phytochemical compounds based on its descriptors. The learned patterns and correlations from the training data were used by the model to make predictions about whether the compounds have inhibitory potential.

F. Mechanism of MACCS analysis

The data used in this instance for the MACCS analysis came from the ChEMBL database, a helpful resource for drug discovery research. Sorting and getting the data ready for more analysis were part of purifying the ChEMBL data. One of the initial stages of the investigation was classifying the data according to the IC50 value. The half-maximal inhibitory concentration (IC50), which quantifies a drug's efficiency in inhibiting a specific biological target, is used in research. The typical IC50 values utilised in this case to classify the samples ranged from 2.5 to 5 ng.

These criteria allowed the chemicals in the dataset to be categorised as either active or inactive drugs. Depending on whether a drug's IC50 value fell inside or beyond the defined range (2.5 to 5 ng), it was classified as either active or inactive. For the MACCS study, the molecular structures of the ligands were designed. This effort was aided by the RDKit module, a powerful Python cheminformatics toolkit. One of the numerous capabilities and tools that RDKit provides for working with molecular structures is the capacity to generate and edit chemical structures. [57] [[35], [58], https://www.rdkit.org"

The RDKit tool was used to generate and show the ligands' molecular structures in a manner suitable for MACCS analysis. This procedure is crucial because it underpins MACCS fingerprints, which are generated from molecular structures and capture important structural components that affect the medication's ability to work.

Following the creation of the molecular fingerprints, the data was processed using a variety of machine learning models, including BaggingClassifier, HistGradientBoostingClassifier, GradientBoostingClassifier, ExtraTreesRegressor,

RandomForestClassifier, AdaBoostClassifier, NuSVC, and SVC. These models were chosen because they could perform classification tasks and were appropriate for the dataset that was provided.

To evaluate the effectiveness of each machine learning model, cross-validation scores were computed. Cross-validation is a method for evaluating the model's performance by repeatedly training and testing the model on different combinations of the dataset's subsets. This approach provides a more trustworthy evaluation of the model's accuracy.

The RandomForestClassifier, which achieved an accuracy of 80%, had the highest efficiency of all the machine learning models examined. With the RandomForestClassifier ensemble learning approach, predictions are made using a combination of several decision trees. It uses the concept of bagging and random feature selection to create a range of powerful models.

After selecting the most useful model, the following stage was to identify the crucial components of the categorisation. The feature_importances function of the RandomForestClassifier was employed in this process. This function quantifies how important each characteristic is in relation to the other throughout the classification process. The graph that follows lists the 40 most important characteristics.

After using these 40 phytochemical characteristics for training, the Machine model performed 89% efficiently, demonstrating the importance of these features to the effectiveness and potential of medications. Through MACCS analysis of major RandomForestClassifier properties, important knowledge for drug development and repurposing is obtained. Researchers can concentrate on certain characteristics that have major effects when they are able to identify important chemical descriptors and substructures. This focused strategy improves the choice of currently available medications for repurposing, saving time and money.

Understanding of key features helps in generating more effective drug candidates by adding or modifying key chemical descriptors and substructures. Utilising biological or disease-specific targets, this optimisation strategy expedites the medication development process.

To identify probable negative reactions and unintended effects, it helps to understand important traits. Medicine chemists can improve safety profiles and maximise medication prospects by identifying specific chemical properties that lead to unfavourable side effects or toxicity. Additionally, comprehending key characteristics helps us grasp how drugs work mechanistically. Structure-function links are clarified by correlating particular molecular descriptors or substructures with categorization results. The effectiveness of medications is improved by using this information to create reasonable compounds.

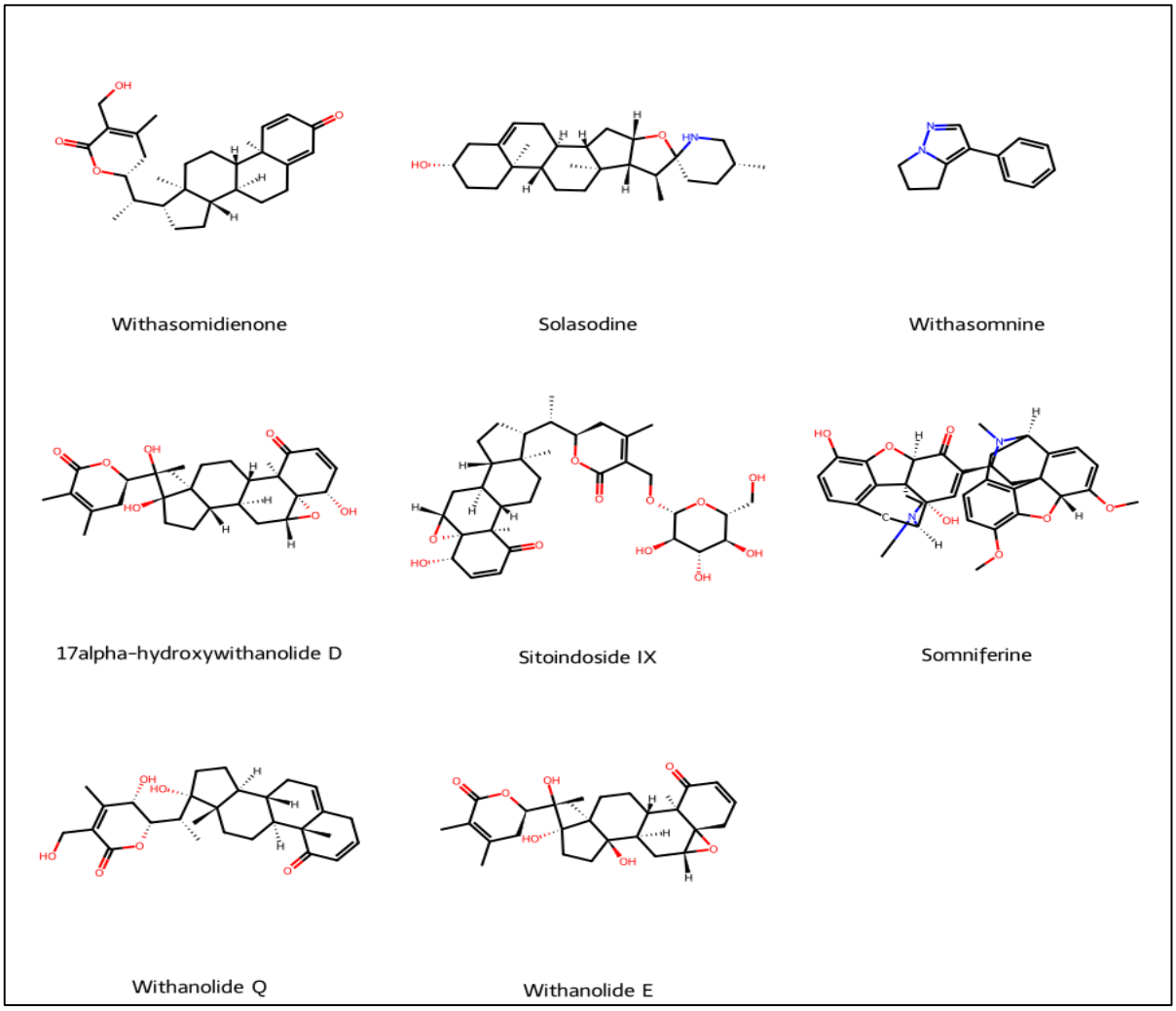


Figure 3 : Molecular structures of phytochemicals of *Diospyros kaki*

CHAPTER 6: RESULTS

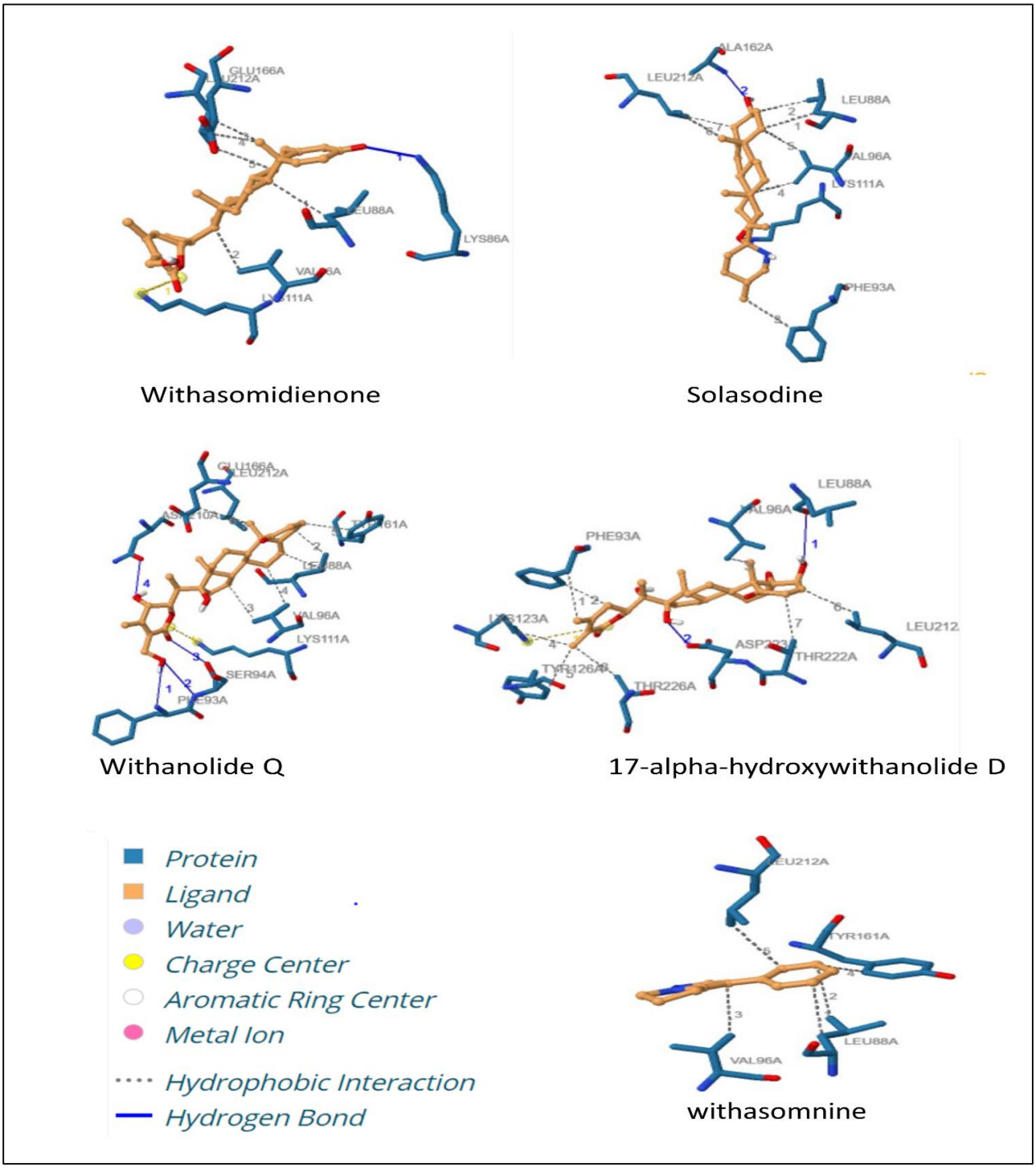


Figure 4: Molecular docking of ligands and target protein (PDK-1) retrieved from PLI

The binding affinity of numerous phytochemicals, including Solasodine, 17-alpha-hydroxywithanolide D, Withanolide Q, Withasomidienone, Withasomnine, and Anahygrine to a particular target of interest was investigated in this work using a docking approach. These phytochemicals acquired binding energies were found to be 11.4 kcal/mol, 10.4 kcal/mol, 9.8 kcal/mol, 7.6 kcal/mol, and 6.7 kcal/mol, respectively. Each phytochemical's interactions with the target protein are indicated by the binding energies that were obtained. A better binding affinity and a greater chance of generating stable complexes are often suggested by lower binding energies.

Table III: Molecular docking of phytochemicals against target protein (PDK-1) in-silico.

Phytochemical	Binding energy(kcal/mol)	No. of bonds	Amino acid participating
Solasodine	-11.4	4	LEU88, VAL96, LEU212, ALA162
17-alpha-hydroxywithanolide D	-10.4	4	VAL96, LEU212, THR222, LEU88
Compound 8i1 (Standard inhibitor)	-10.1	9	ALA 162, ASN 210, ASP 223
Withanolide Q	-9.8	5	LEU88, VAL96, TYR161, GLU166, LEU212
Withasomidienone	-9.6	3	LEU88, VAL96, LEU212
Withasomnine	-7.1	4	LEU88, VAL96, TYR161, LEU212
Anahygrine	-6.7	4	LEU88, VAL96, THR222, ALA109

Following a successful docking analysis, this study evaluated the medications further to determine their pharmacodynamics and pharmacokinetics characteristics utilising bioinformatics software tools including swisADME and Molinspiration.In accordance with the Lipinski rule of five, which assesses drug-likeness based on physicochemical features, 17-alpha-hydroxywithanolide D, Withanolide Q, Withanolide E, and Withasomnine exhibited no violations. The Lipinski rule of five was shown to be broken by Withasomidienone, Solasodine, Somniferine, and Sitoindoside IX, suggesting possible difficulties in their absorption, distribution, metabolism, and excretion. Further

SwisADME analysis revealed that Withasomidienone, Solasodine, and Withasomnine can cross the blood-brain barrier. These phytochemicals show potential as effective

inhibitors of PDK-1 in glioblastoma cancer, highlighting their significance for central.

Table IV: Lipinski's Rule of Five Analysis (RO5) Mass<500, H-bond donors<5, Hbond acceptors<10, LogP value<5

Phytochemicals	Mass g/mol	Hydrogen bond donor	Hydrogen bond acceptor	LogP value	Molar refractivity	No. of violations
Withasomidienone	436.8	1	4	4.31	126	1
Solasodine	413.64	2	3	4.9	127	1
Withasomnine	184.24	0	1	2.34	56.56	0
17alpha-hydroxywithanolide D	486.60	3	7	1.95	128.73	0
Sitoindoside IX	632.74	5	11	0.48	156.87	2
Somniferine	608.68	2	9	1.67	171	1
Withanolide Q	470.67	3	6	2.67	129	0
Withanolide E	486.60	3	7	1.95	128	0

In addition, the MOLinspiration software's pharmacokinetics analysis included a number of aspects, including GPCR, kinase inhibitor, enzyme inhibitor, NRL (Nuclear Receptor Ligand), and protease inhibitor. These investigations shed light on the inhibitory effect of the medications as well as any potential interactions between them and certain biological targets. Withanolide Q, Solasodine, 17-alpha-hydroxywithanolide D, Withasomidienone, and Withasomnine's pharmacokinetics and pharmacodynamics are all discussed in detail in the generated results from these bioinformatics analysis. Understanding the possible effectiveness and safety characteristics of these compounds thanks to this thorough assessment will help direct future research and development initiatives in medication discovery and optimisation.

Table V: Bioactivity Scores of Phytochemicals retrieved from Molinspiration.

Phytochemical	GPCR	PI	EI	NRL	KI	ICM
Withasomidienone	0.04	0.07	0.73	0.96	-0.66	-0.03
Solasodine	0.24	0.01	0.60	0.36	-0.66	-0.17
Withasomnine	-0.49	-0.43	0.58	-0.10	0.58	-0.43
17alpha-hydroxywithanolide D	0.06	0.18	1.01	0.78	-0.43	0.20
Sitoindoside IX	-0.08	0.11	0.47	0.06	-0.77	-0.49
Somniferine	0.69	-0.05	-0.16	-0.41	-0.53	-0.57

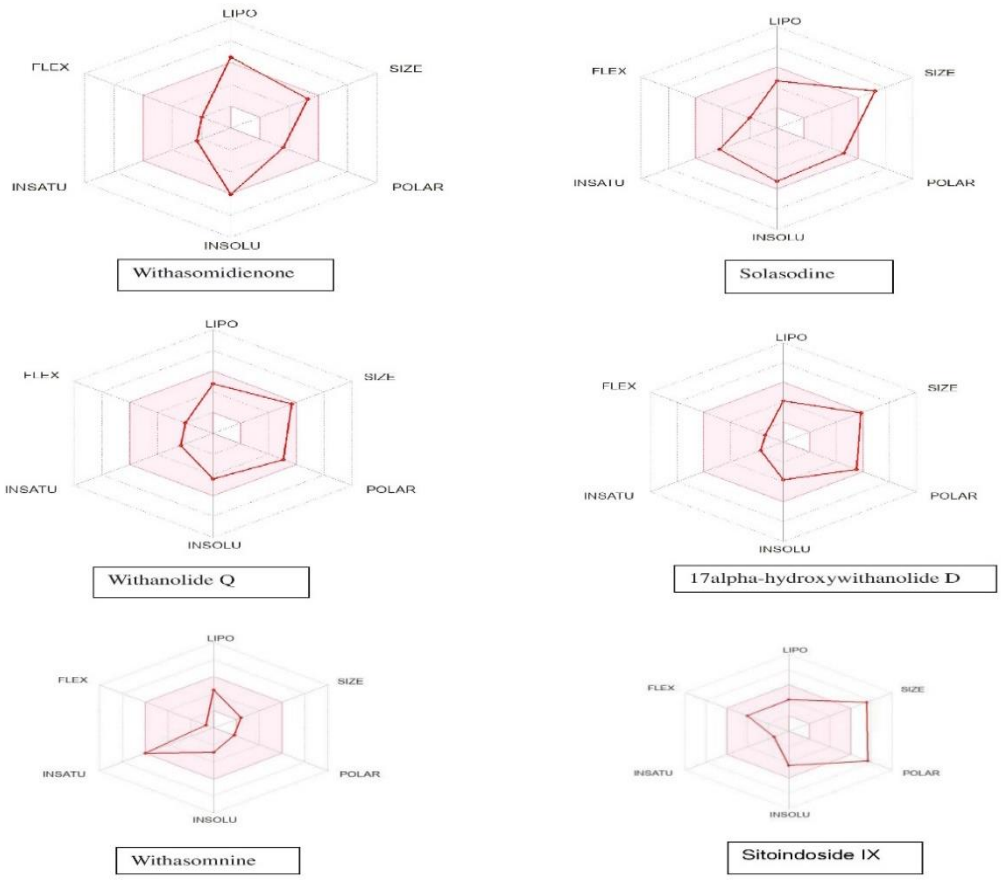
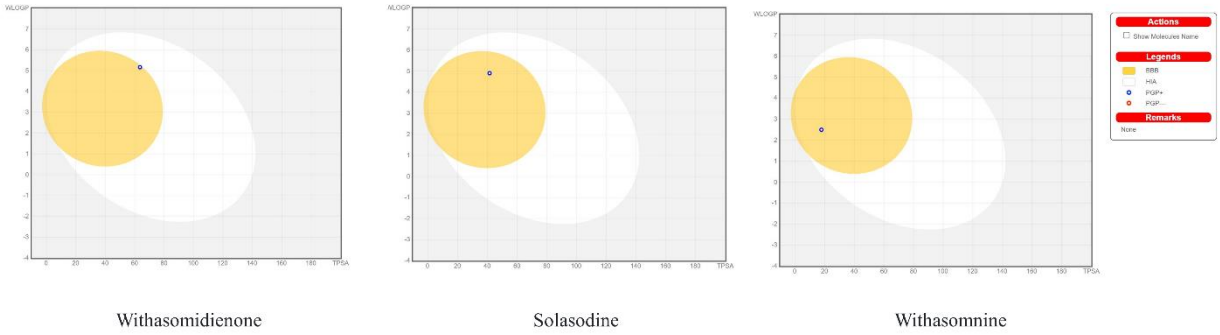


Figure 5: Bioavailability radars of retrieved using swisADME



Withasomidienone Solasodine Withasomnine

Figure 6: SwisADME analysis reveals phytochemicals Withasomidienone, Solasodine, Withasomnine Crossing blood-brain barrier, unlocking therapeutic potential for brain health.

A machine learning study was done on the chosen phytochemicals to verify our docking results. The RandomForestClassifier was chosen among the many machine learning models because it demonstrated the best accuracy and F1 value, showing its usefulness in predicting the activity of the chemicals. The accuracy of the Machine learning model came out to be 80%.

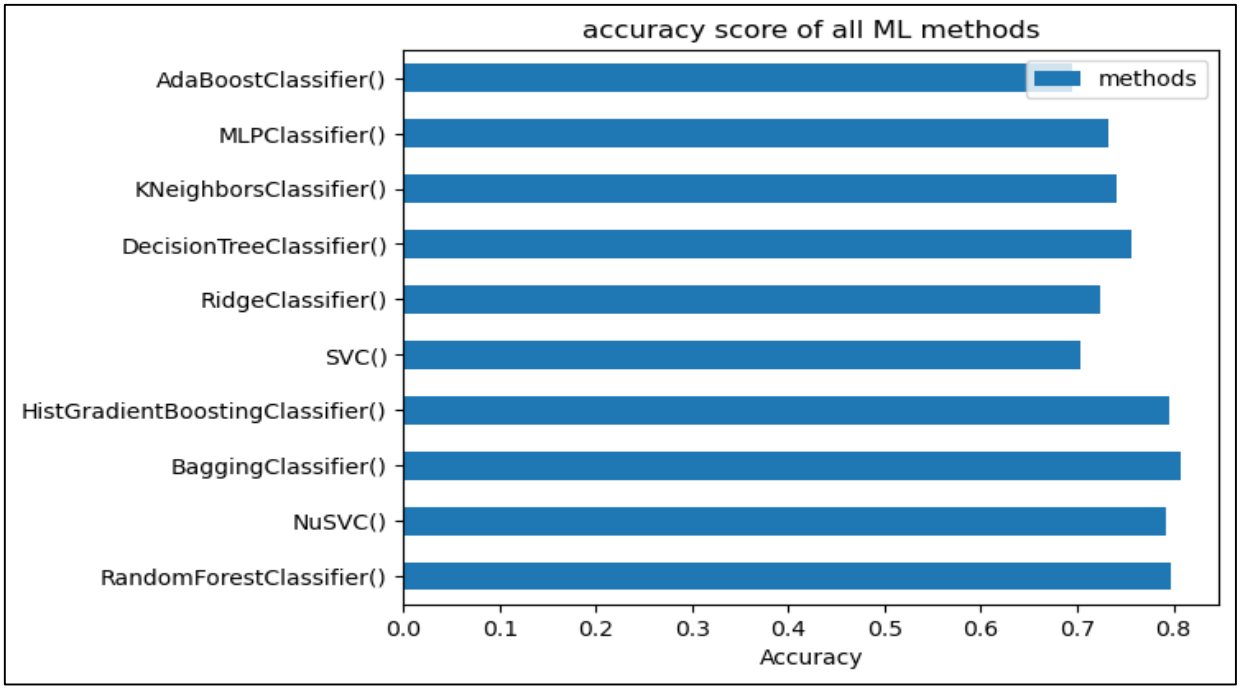


Figure 7: Bar graph Depicting different machine learning model and their accuracy

Confusion matrix analysis has been used to evaluate the performance of a classification model. It provides a summary of the model's predictions compared to the actual outcomes.

A confusion matrix is a table that displays the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for a binary classification problem. The rows of the matrix represent the actual classes, while the columns represent the predicted classes.

The confusion matrix allows us to calculate various performance metrics, including accuracy, precision, recall, and F1 score, which provide insights into different aspects of the model's performance.

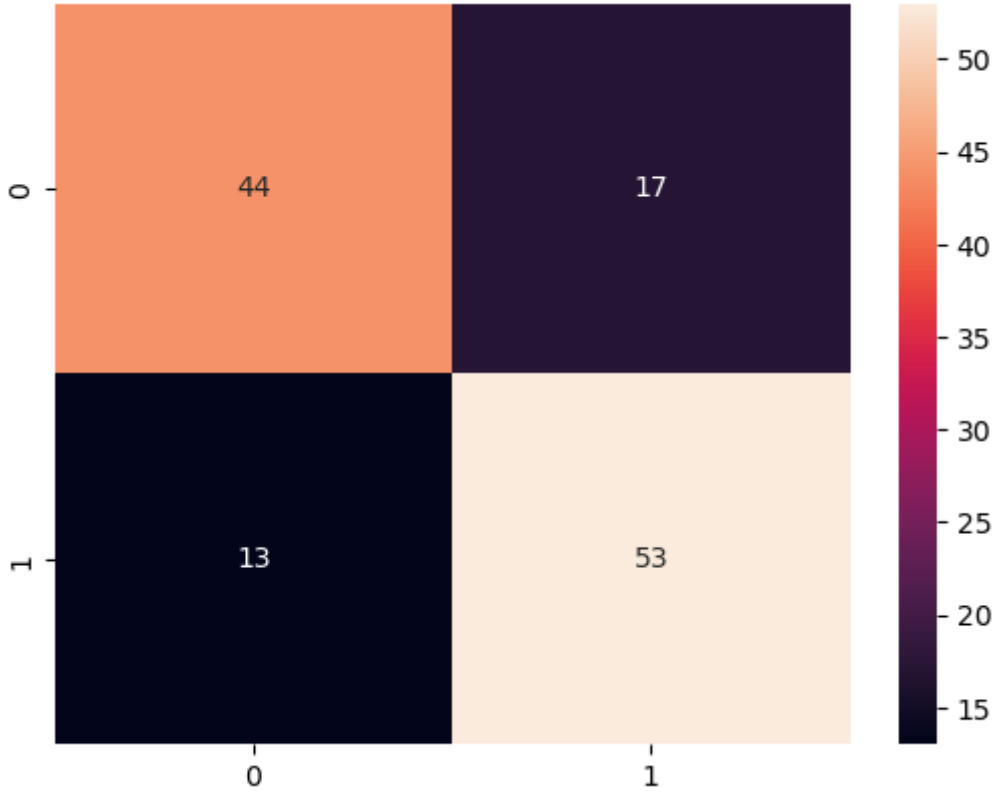


Figure 8: Confusion matrix of Machine Learning accuracy. Depicting True Positive=46, True Negative= 56, False Positive= 16, False Negative=9.

The research revealed six phytochemicals as possible active medicines for inhibiting the PDK-1 protein using the RandomForestClassifier model. Solasodine, 17-alpha-hydroxywithanolide D, withanolide Q, withasomidienone, and withasomnine were these phytochemicals. The machine learning algorithm made predictions about the compounds' possible inhibitory action against the target protein using a variety of the compounds' properties and traits. This machine learning-based method offers insightful information about the potential therapeutic use of these phytochemicals in PDK-1 protein targeting. The medicines that were shown to be active may make excellent candidates for additional experimental testing and development as possible PDK-1 inhibitors. This protein is very relevant to many biological processes and disease pathways.

In order to optimise and reconstruct pharmaceuticals, we examined the MACCS characteristics of active medications with consistent IC50 values. A set of 166 specified substructural patterns serve as the foundation for MACCS descriptors, which are binary fingerprints that encode structural information. Specific pieces or functional groupings are either present or absent in these patterns. We discovered common structural traits among the active medicines by looking at the MACCS descriptors. The ability to discover important molecular properties that can be improved makes this knowledge useful for changing and improving medications. Through the use of the MACCS descriptors, we may identify connections between therapeutic action and molecular structure that can inform future drug design changes and optimisation techniques.

Table VI: List of MAACS fingerprint keys

MDL MACCS Key	**Smart Pattern**	42	Aliphatic secondary nitrogen
1	Aromatic ring	43	Aromatic primary nitrogen
2	Aliphatic ring	44	Aliphatic primary nitrogen
3	Double bond	45	Aromatic halogen
4	Triple bond	46	Aliphatic halogen
5	Aromatic nitrogen	47	Aromatic ester
6	Aliphatic nitrogen	48	Aliphatic ester
7	Aromatic oxygen	49	Aromatic ether
8	Aliphatic oxygen	50	Aliphatic ether
9	Aromatic sulfur	51	Aromatic amide
10	Aliphatic sulfur	52	Aliphatic amide
11	Aromatic chlorine	53	Aromatic nitrile
12	Aliphatic chlorine	54	Aliphatic nitrile
13	Aromatic bromine	55	Aromatic urea
14	Aliphatic bromine	56	Aliphatic urea
15	Aromatic iodine	57	Aromatic thioether
16	Aliphatic iodine	58	Aliphatic thioether
17	Aromatic hydroxyl	59	Aromatic imide
18	Aliphatic hydroxyl	60	Aliphatic imide
19	Aromatic methoxy	50	Aliphatic ether
20	Aliphatic methoxy	51	Aromatic amide
21	Aromatic amino	52	Aliphatic amide
22	Aliphatic amino	53	Aromatic nitrile
23	Aromatic thiol	54	Aliphatic nitrile
24	Aliphatic thiol	55	Aromatic urea
25	Aromatic nitro	56	Aliphatic urea
26	Aliphatic nitro	57	Aromatic thioether
27	Aromatic carbonyl	58	Aliphatic thioether
28	Aliphatic carbonyl	59	Aromatic imide
29	Aromatic carboxylic acid	60	Aliphatic imide

30	Aliphatic carboxylic acid	56	Aliphatic urea
31	Aromatic sulfonic acid	57	Aromatic thioether
32	Aliphatic sulfonic acid	58	Aliphatic thioether
33	Aromatic phosphonic acid	59	Aromatic imide
34	Aliphatic phosphonic acid	60	Aliphatic imide
35	Aromatic phosphinic acid	61	Aromatic hydrazine
36	Aliphatic phosphinic acid	62	Aliphatic hydrazine
37	Aromatic quaternary nitrogen	63	Aromatic azomethine
38	Aliphatic quaternary nitrogen	64	Aliphatic azomethine
39	Aromatic tertiary nitrogen	65	Aromatic azo
40	Aliphatic tertiary nitrogen	66	Aliphatic azo
41	Aromatic secondary nitrogen	67	Aromatic Schiff base
68	Aliphatic Schiff base	111	Aromatic acenaphthylene
69	Aromatic pyrazole	112	Aliphatic acenaphthylene
70	Aliphatic pyrazole	113	Aromatic acenaphthene
71	Aromatic imidazole	114	Aliphatic acenaphthene
72	Aliphatic imidazole	115	Aromatic fluorene
73	Aromatic thiazole	116	Aliphatic fluorene
74	Aliphatic thiazole	117	Aromatic phenanthrene
75	Aromatic furan	118	Aliphatic phenanthrene
76	Aliphatic furan	119	Aromatic anthracene
77	Aromatic pyrrole		
78	Aliphatic pyrrole	120	Aliphatic anthracene
79	Aromatic pyridine	121	Aromatic pyrene
80	Aliphatic pyridine	122	Aliphatic pyrene

81	Aromatic quinoline	123	Aromatic benz[a]anthracene
82	Aliphatic quinoline	124	Aliphatic benz[a]anthracene
83	Aromatic isoquinoline	125	Aromatic chrysene
84	Aliphatic isoquinoline	126	Aliphatic chrysene
85	Aromatic carbazole	127	Aromatic benzo[a]pyrene
86	Aliphatic carbazole	128	Aliphatic benzo[a]pyrene
87	Aromatic phthalazine	129	Aromatic benzo[b]fluoranthene
88	Aliphatic phthalazine	130	Aliphatic benzo[b]fluoranthene
89	Aromatic phenanthrene	131	Aromatic benzo[k]fluoranthene
90	Aliphatic phenanthrene	132	Aliphatic benzo[k]fluoranthene
91	Aromatic anthracene	133	Aromatic dibenz[a,h]anthracene
92	Aliphatic anthracene	134	Aliphatic dibenz[a,h]anthracene
93	Aromatic phenanthrene	135	Aromatic benzo[ghi]perylene
94	Aliphatic phenanthrene	136	Aliphatic benzo[ghi]perylene
95	Aromatic chrysene	137	Aromatic indeno[1,2,3-cd]pyrene
96	Aliphatic chrysene	138	Aliphatic indeno[1,2,3-cd]pyrene
97	Aromatic benzo[a]pyrene	139	Aromatic naphthalene
98	Aliphatic benzo[a]pyrene	140	Aliphatic naphthalene
99	Aromatic benzo[b]fluoranthene	141	Aromatic acenaphthylene
100	Aliphatic benzo[b]fluoranthene	142	Aliphatic acenaphthylene

101	Aromatic benzo[k]fluoranthene	143	Aromatic acenaphthene
102	Aliphatic benzo[k]fluoranthene	144	Aliphatic acenaphthene
103	Aromatic dibenz[a,h]anthracene	145	Aromatic fluorene
104	Aliphatic dibenz[a,h]anthracene	146	Aliphatic fluorene
105	Aromatic benzo[ghi]perylene	147	Aromatic phenanthrene
106	Aliphatic benzo[ghi]perylene	148	Aliphatic phenanthrene
107	Aromatic indeno[1,2,3-cd]pyrene	149	Aromatic anthracene
108	Aliphatic indeno[1,2,3-cd]pyrene	150	Aliphatic anthracene
109	Aromatic naphthalene	151	Aromatic pyrene
110	Aliphatic naphthalene	152	Aliphatic pyrene
157	Aromatic benzo[a]pyrene	162	Aliphatic acenaphthene
158	Aliphatic naphthalene	163	Aromatic fluorene
159	Aromatic acenaphthylene	164	Aliphatic fluorene
160	Aliphatic acenaphthylene	165	Aromatic phenanthrene
161	Aromatic acenaphthene	166	Aliphatic phenanthrene

The binding affinity of numerous phytochemicals, including Solasodine, 17-alpha-hydroxywithanolide D, Withanolide Q, Withasomidienone, Withasomnine, and Anahygrine to a particular target of interest was investigated in this work using a docking approach. These phytochemicals' acquired binding energies were found to be 11.4 kcal/mol, 10.4 kcal/mol, 9.8 kcal/mol, 7.6 kcal/mol, and 6.7 kcal/mol, respectively. Each phytochemical's interactions with the target protein are indicated by the binding energies that were obtained. A better binding affinity and a greater chance of generating stable complexes are often suggested by lower binding energies.

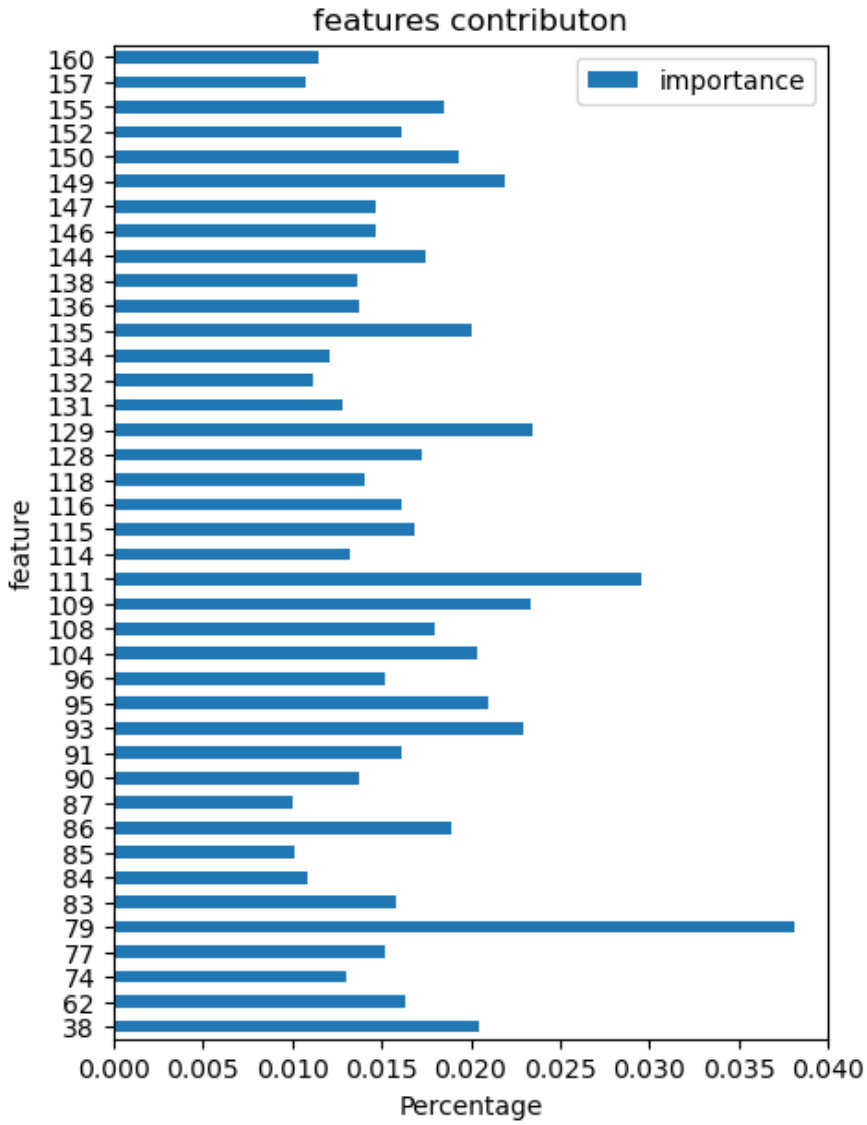


Figure 9: Contribution of each key based on their presence in the drug/compound.

The intersection of molecular descriptors present in all the drugs includes descriptors 79, 86, 95, 111, 38, 91, 109, 128, 135, 149, 150, 62, and more. These descriptors indicate the presence of specific chemical structures or functional groups in the active drug molecules. For instance, aromatic pyridine, aliphatic carbazole, aromatic chrysene, aromatic acenaphthylene, aliphatic quaternary nitrogen, aromatic anthracene, aromatic naphthalene, aliphatic benzopyrene, aromatic benzo perylene, aliphatic anthracene, aliphatic hydrazine, and others.

The consistent presence of these descriptors suggests that these structural features, such as aromatic rings or specific functional groups, play a significant role in the activity of the drugs. These findings provide valuable insights into the importance of certain chemical structures or fragments for the activity of our active drugs and can guide further drug design and optimization strategies.

CHAPTER 7: CONCLUSION

The adverse side effects and toxicities associated with synthetic drugs commonly used in cancer treatment have prompted a shift in focus towards exploring alternative therapeutic agents that are both non-toxic and readily available. In this context, phytochemicals derived from natural sources have gained attention as potential anti-cancer agents.

Withania somnifera, a plant known for its various health benefits, possesses a diverse array of bioactive compounds within its phytochemical profile. Therefore, a specific research endeavor was conducted using computational methods (in-silico study) to investigate the potential of utilizing bioactive compounds derived from *Withania somnifera* to suppress the activity of a protein called PDK-1. PDK-1 plays a crucial role in the AKT-mTOR pathway, which is involved in cancer progression.

In addition to targeting PDK-1, the we considered several additional parameters to evaluate the suitability of phytochemicals as potential therapeutic drugs. These parameters encompassed minimal cytotoxicity (toxic effects on cells), maximum bioavailability (ability to reach the target site in the body), bioactivity (capability to induce desired biological effects), and blood-brain barrier (BBB) permeation (ability to cross the protective barrier surrounding the brain). By considering these factors, the we categorized the phytochemicals from *Withania somnifera* based on their potential as therapeutic drugs. Further MACCS analysis of the active compound demonstrated that phytochemicals exhibit drug activity-associated fingerprints. Consequently, slight structural modifications of phytochemicals hold the potential to enhance their efficacy as potent drugs or inhibitors specifically targeting glioblastoma.

The study also suggests that the approach of inhibiting other proteins involved in the AKT-mTOR pathway could be employed, implying a broader application of the research findings. Furthermore, the statement emphasizes the importance of further research in this field to enhance our understanding of the underlying molecular mechanisms responsible for the recurrence and persistence of glioblastoma stem cells (GSCs). By gaining more insights, we can precisely identify the specific proteins involved, potentially leading to the development of more targeted and effective therapeutic strategies for GSC-related cancers.

REFERENCES

[1] S. Daniele *et al.*, "Dual Inhibition of PDK1 and Aurora Kinase A: An Effective Strategy to Induce Differentiation and Apoptosis of Human Glioblastoma Multiforme Stem Cells," *ACS Chem Neurosci*, vol. 8, no. 1, pp. 100–114, Jan. 2017, doi: 10.1021/acschemneuro.6b00251.

[2] E. Majewska, "AKT/GSK3β Signaling in Glioblastoma," 2044, doi: 10.1007/s11064-016-2044-x/Published.

[3] M. Signore *et al.*, "Combined PDK1 and CHK1 inhibition is required to kill glioblastoma stem-like cells in vitro and in vivo," *Cell Death Dis*, vol. 5, no. 5, 2014, doi: 10.1038/cddis.2014.188.

[4] H. Xu *et al.*, "Epidermal growth factor receptor in glioblastoma (Review)," *Oncol Lett*, vol. 14, no. 1, pp. 512–516, 2017, doi: 10.3892/ol.2017.6221.

[5] H. J. Lin, F. C. Hsieh, H. Song, and J. Lin, "Elevated phosphorylation and activation of PDK-1/AKT pathway in human breast cancer," *Br J Cancer*, vol. 93, no. 12, pp. 1372–1381, Dec. 2005, doi: 10.1038/sj.bjc.6602862.

[6] L. E. Mendie and S. Hemalatha, "Molecular Docking of Phytochemicals Targeting GFRs as Therapeutic Sites for Cancer: an In Silico Study," *Appl Biochem Biotechnol*, vol. 194, no. 1, pp. 215–231, Jan. 2022, doi: 10.1007/s12010-021-03791-7.

[7] R. Direito, J. Rocha, B. Sepodes, and M. Eduardo-Figueira, "From withania somnifera L. (persimmon) phytochemical profile and health impact to new product perspectives and waste valorization," *Nutrients*, vol. 13, no. 9. MDPI, Sep. 01, 2021. doi: 10.3390/nu13093283.

[8] C. W. Yap, "PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints," *J Comput Chem*, vol. 32, no. 7, pp. 1466–1474, May 2011, doi: 10.1002/jcc.21707.

[9] C. W. Yap, "PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints," *J Comput Chem*, vol. 32, no. 7, pp. 1466–1474, May 2011, doi: 10.1002/jcc.21707.

[10] H. Moriwaki, Y. S. Tian, N. Kawashita, and T. Takagi, "Mordred: A molecular descriptor calculator," *J Cheminform*, vol. 10, no. 1, Feb. 2018, doi: 10.1186/s13321-018-0258-y.

[11] G. Ramos *et al.*, "Fatigue Evaluation through Machine Learning and a Global Fatigue Descriptor," *J Healthc Eng*, vol. 2020, 2020, doi: 10.1155/2020/6484129.

[12] S. S. S. J. Ahmed and V. Ramakrishnan, "Systems biological approach of molecular descriptors connectivity: Optimal descriptors for oral bioavailability prediction," *PLoS One*, vol. 7, no. 7, Jul. 2012, doi: 10.1371/journal.pone.0040654.

[13] A. C. Tan, D. M. Ashley, G. Y. López, M. Malinzak, H. S. Friedman, and M. Khasraw, "Management of glioblastoma: State of the art and future directions," *CA Cancer J Clin*, vol. 70, no. 4, pp. 299–312, Jul. 2020, doi: 10.3322/caac.21613.

[14] A. R. P. Antunes, I. Scheyltjens, J. Duerinck, B. Neyns, K. Movahedi, and J. A. Van Ginderachter, "Understanding the glioblastoma immune microenvironment as basis for the development of new immunotherapeutic strategies," *Elife*, vol. 9, Feb. 2020, doi: 10.7554/eLife.52176.

[15] T. Hara *et al.*, "Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma," *Cancer Cell*, vol. 39, no. 6, pp. 779-792.e11, Jun. 2021, doi: 10.1016/j.ccell.2021.05.002.

[16] S. Daisy Precilla, I. Biswas, S. S. Kuduvalli, and T. S. Anitha, "Crosstalk between PI3K/AKT/mTOR and WNT/β-Catenin signaling in GBM - Could combination therapy checkmate the collusion?," *Cellular Signalling*, vol. 95. Elsevier Inc., Jul. 01, 2022. doi: 10.1016/j.cellsig.2022.110350.

[17] H. Kim *et al.*, "Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution," *Genome Res*, vol. 25, no. 3, pp. 316–327, Mar. 2015, doi: 10.1101/gr.180612.114.

[18] C. Neftel *et al.*, "An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma," *Cell*, vol. 178, no. 4, pp. 835-849.e21, Aug. 2019, doi: 10.1016/j.cell.2019.06.024.

[19] N. Montemurro, "Glioblastoma Multiforme and Genetic Mutations: The Issue Is Not over YetAn Overview of the Current Literature," *Journal of Neurological Surgery, Part A: Central European Neurosurgery*, vol. 81, no. 1. Georg Thieme Verlag, pp. 64–70, 2020. doi: 10.1055/s-0039-1688911.

[20] R. McLendon *et al.*, "Comprehensive genomic characterization defines human glioblastoma genes and core pathways," *Nature*, vol. 455, no. 7216, pp. 1061–1068, Oct. 2008, doi: 10.1038/nature07385.

[21] S. Han *et al.*, "IDH mutation in glioma: molecular mechanisms and potential therapeutic targets," *British Journal of Cancer*, vol. 122, no. 11. Springer Nature, pp. 1580–1589, May 26, 2020. doi: 10.1038/s41416-020-0814-x.

[22] F. Han *et al.*, "PTEN gene mutations correlate to poor prognosis in glioma patients: A meta-analysis," *Onco Targets Ther*, vol. 9, pp. 3485–3492, Jun. 2016, doi: 10.2147/OTT.S99942.

[23] M. Simon *et al.*, "TERT promoter mutations: A novel independent prognostic factor in primary glioblastomas," *Neuro Oncol*, vol. 17, no. 1, pp. 45–52, Jan. 2015, doi: 10.1093/neuonc/nou158.

[24] H. Colman, "Toward more informative biomarker-based clinical trials in glioblastoma," *Neuro-Oncology*, vol. 19, no. 7. Oxford University Press, pp. 880–881, Jul. 01, 2017. doi: 10.1093/neuonc/nox063.

[25] Q. Sun *et al.*, "Up-regulation of MSH6 is associated with temozolomide resistance in human glioblastoma," *Biochem Biophys Res Commun*, vol. 496, no. 4, pp. 1040–1046, Feb. 2018, doi: 10.1016/j.bbrc.2018.01.093.

[26] A. B. Behrooz, Z. Talaie, F. Jusheghani, M. J. Łos, T. Klonisch, and S. Ghavami, "Wnt and PI3K/Akt/mTOR Survival Pathways as Therapeutic Targets in Glioblastoma,"

International Journal of Molecular Sciences, vol. 23, no. 3. MDPI, Feb. 01, 2022. doi: 10.3390/ijms23031353.

[27] X. Li *et al.*, "PI3K/Akt/mTOR signaling pathway and targeted therapy for glioblastoma," 2016. [Online]. Available: www.impactjournals.com/oncotarget

[28] K. K. umar Velpula and A. J. Tsung, "PDK1: a new therapeutic target for glioblastoma?," *CNS oncology*, vol. 3, no. 3. pp. 177–179, May 01, 2014. doi: 10.2217/cns.14.13.

[29] G. Hibino, T. Nadamoto, F. Fujisawa, and T. Fushiki, "Regulation of the Peripheral Body Temperature by Foods: A Temperature Decrease Induced by the Japanese Persimmon (kaki, Withania somnifera)," *Biosci Biotechnol Biochem*, vol. 67, no. 1, pp. 23–28, Jan. 2003, doi: 10.1271/bbb.67.23.

[30] M. S. Butt *et al.*, "Persimmon (withania somnifera) fruit: Hidden phytochemicals and health claims," *EXCLI Journal*, vol. 14. Leibniz Research Centre for Working Environment and Human Factors, pp. 542–561, May 04, 2015. doi: 10.17179/excli2015-159.

[31] P. H. M. Torres, A. C. R. Sodero, P. Jofily, and F. P. Silva-Jr, "Key topics in molecular docking for drug design," *International Journal of Molecular Sciences*, vol. 20, no. 18. MDPI AG, Sep. 02, 2019. doi: 10.3390/ijms20184574.

[32] L. Pinzi and G. Rastelli, "Molecular docking: Shifting paradigms in drug discovery," *International Journal of Molecular Sciences*, vol. 20, no. 18. MDPI AG, Sep. 01, 2019. doi: 10.3390/ijms20184331.

[33] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *J Comput Chem*, p. NA-NA, 2009, doi: 10.1002/jcc.21334.

[34] T. Nittoli *et al.*, "The identification of 8,9-dimethoxy-5-(2-aminoalkoxy-pyridin-3-yl)-benzo[c][2,7]naphthyridin-4-ylamines as potent inhibitors of 3-phosphoinositide-dependent kinase-1 (PDK-1)," *Eur J Med Chem*, vol. 45, no. 4, pp. 1379–1386, Apr. 2010, doi: 10.1016/j.ejmech.2009.12.036.

[35] A. Korotcov, V. Tkachenko, D. P. Russo, and S. Ekins, "Comparison of Deep Learning with Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets," *Mol Pharm*, vol. 14, no. 12, pp. 4462–4475, Dec. 2017, doi: 10.1021/acs.molpharmaceut.7b00578.

[36] A. Gaulton *et al.*, "The ChEMBL database in 2017," *Nucleic Acids Res*, vol. 45, no. D1, pp. D945–D954, Jan. 2017, doi: 10.1093/nar/gkw1074.

[37] M. T. Patrick *et al.*, "Drug Repurposing Prediction for Immune-Mediated Cutaneous Diseases using a Word-Embedding–Based Machine Learning Approach," *Journal of Investigative Dermatology*, vol. 139, no. 3, pp. 683–691, Mar. 2019, doi: 10.1016/j.jid.2018.09.018.

[38] J. L. Durant, B. A. Leland, D. R. Henry, and J. G. Nourse, "Reoptimization of MDL keys for use in drug discovery," *J Chem Inf Comput Sci*, vol. 42, no. 6, pp. 1273–1280, Nov. 2002, doi: 10.1021/ci010132r.

[39] N. Cerisier, L. Regad, D. Triki, M. Petitjean, D. Flatters, and A. C. Camproux, “Statistical Profiling of One Promiscuous Protein Binding Site: Illustrated by Urokinase Catalytic Domain,” *Mol Inform*, vol. 36, no. 10, Oct. 2017, doi: 10.1002/minf.201700040.

[40] K. Mohanraj *et al.*, “IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry and Therapeutics,” *Sci Rep*, vol. 8, no. 1, Dec. 2018, doi: 10.1038/s41598-018-22631-z.

[41] H. M. Berman *et al.*, “The Protein Data Bank,” 2000. [Online]. Available: http://www.rcsb.org/pdb/status.html

[42] T. Gaillard, “Evaluation of AutoDock and AutoDock Vina on the CASF-2013 Benchmark,” *J Chem Inf Model*, vol. 58, no. 8, pp. 1697–1706, Aug. 2018, doi: 10.1021/acs.jcim.8b00312.

[43] D. Seeliger and B. L. De Groot, “Ligand docking and binding site analysis with PyMOL and Autodock/Vina,” *J Comput Aided Mol Des*, vol. 24, no. 5, pp. 417–422, 2010, doi: 10.1007/s10822-010-9352-6.

[44] N. T. Nguyen *et al.*, “Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity,” *J Chem Inf Model*, vol. 60, no. 1, pp. 204–211, Jan. 2020, doi: 10.1021/acs.jcim.9b00778.

[45] S. Salentin *et al.*, “From malaria to cancer: Computational drug repositioning of amodiaquine using PLIP interaction patterns,” *Sci Rep*, vol. 7, no. 1, Dec. 2017, doi: 10.1038/s41598-017-11924-4.

[46] S. Salentin, S. Schreiber, V. J. Haupt, M. F. Adasme, and M. Schroeder, “PLIP: Fully automated protein-ligand interaction profiler,” *Nucleic Acids Res*, vol. 43, no. W1, pp. W443–W447, 2015, doi: 10.1093/nar/gkv315.

[47] M. F. Adasme *et al.*, “PLIP 2021: Expanding the scope of the protein-ligand interaction profiler to DNA and RNA,” *Nucleic Acids Res*, vol. 49, no. W1, pp. W530–W534, Jul. 2021, doi: 10.1093/nar/gkab294.

[48] “Discovery Studio Life Science Modeling and Simulations.”

[49] D. Mendez *et al.*, “ChEMBL: Towards direct deposition of bioassay data,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D930–D940, Jan. 2019, doi: 10.1093/nar/gky1075.

[50] A. Gaulton *et al.*, “ChEMBL: A large-scale bioactivity database for drug discovery,” *Nucleic Acids Res*, vol. 40, no. D1, Jan. 2012, doi: 10.1093/nar/gkr777.

[51] A. Mayr *et al.*, “Large-scale comparison of machine learning methods for drug target prediction on ChEMBL,” *Chem Sci*, vol. 9, no. 24, pp. 5441–5451, 2018, doi: 10.1039/c8sc00148k.

[52] A. Gaulton *et al.*, “ChEMBL: A large-scale bioactivity database for drug discovery,” *Nucleic Acids Res*, vol. 40, no. D1, Jan. 2012, doi: 10.1093/nar/gkr777.

[53] A. Daina, O. Michielin, and V. Zoete, “SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules,” *Sci Rep*, vol. 7, Mar. 2017, doi: 10.1038/srep42717.

[54] C. A. Lipinski, “Lead- and drug-like compounds: the rule-of-five revolution,” *Drug Discov Today Technol*, vol. 1, no. 4, pp. 337–341, Dec. 2004, doi: 10.1016/j.ddtec.2004.11.007.

[55] C. Sommer and D. W. Gerlich, “Machine learning in cell biology-teaching computers to recognize phenotypes,” *J Cell Sci*, vol. 126, no. 24, pp. 5529–5539, Dec. 2013, doi: 10.1242/jcs.123604.

[56] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, “Machine Learning and Data Mining Methods in Diabetes Research,” *Computational and Structural Biotechnology Journal*, vol. 15. Elsevier B.V., pp. 104–116, 2017. doi: 10.1016/j.csbj.2016.12.005.

[57] R. Rodríguez-Pérez and J. Bajorath, “Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values,” *J Med Chem*, vol. 63, no. 16, pp. 8761–8777, Aug. 2020, doi: 10.1021/acs.jmedchem.9b01101.

[58] T. R. Lane, D. H. Foil, E. Minerali, F. Urbina, K. M. Zorn, and S. Ekins, “Bioactivity Comparison across Multiple Machine Learning Algorithms Using over 5000 Datasets for Drug Discovery,” *Mol Pharm*, vol. 18, no. 1, pp. 403–415, Jan. 2021, doi: 10.1021/acs.molpharmaceut.0c01013.

turnitin

Similarity Report ID: oid:27535:36266412

9% Overall Similarity

Top sources found in the following databases:

- 6% Internet database
- Crossref database
- 7% Submitted Works database
- 5% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	dspace.dtu.ac.in:8080 Internet	2%
2	dataset.drugtargetcommons.org Internet	<1%
3	Huma Aslam Butt, Hina Aslam Butt, Arif-ullah Khan. "In Silico Analysis ... Crossref	<1%
4	mdpi.com Internet	<1%
5	link.springer.com Internet	<1%
6	nature.com Internet	<1%
7	Technical University of Cluj-Napoca on 2022-08-31 Submitted works	<1%
8	Staffordshire University on 2020-11-26 Submitted works	<1%

turnitin **Similarity Report ID**: oid:27535:36266412

9	CSU, San Jose State University on 2022-12-12 Submitted works	<1%
10	Colorado School of Mines on 2022-04-08 Submitted works	<1%
11	academic.oup.com Internet	<1%
12	hindawi.com Internet	<1%
13	ncbi.nlm.nih.gov Internet	<1%
14	swissmodel.expasy.org Internet	<1%
15	University of Maryland, University College on 2013-11-10 Submitted works	<1%
16	ascopubs.org Internet	<1%
17	Harrisburg University of Science and Technology on 2022-04-13 Submitted works	<1%
18	University College London on 2023-05-12 Submitted works	<1%
19	Mohamed Diwan M. AbdulHameed, Adel Hamza, Chang-Guo Zhan. "Mi... Crossref	<1%
20	rcastoragev2.blob.core.windows.net Internet	<1%

turnitin Similarity Report ID: oid:27535:36266412

21	shodhbhagirathi.iitr.ac.in:8081 Internet	<1%
22	Kingston University on 2023-05-01 Submitted works	<1%
23	Sandra Ghayad, Pascale Cohen. "Inhibitors of the PI3K/Akt/mTOR Path... Crossref	<1%
24	Sim University on 2011-08-29 Submitted works	<1%
25	University of Leeds on 2018-05-01 Submitted works	<1%
26	"Handbook of Chemoinformatics", Wiley, 2003 Crossref	<1%
27	biorxiv.org Internet	<1%
28	frontiersin.org Internet	<1%
29	Indian Institute of Management, Indore on 2020-06-19 Submitted works	<1%
30	Jenny Balfer, Jürgen Bajorath. "Introduction of a Methodology for Visu... Crossref	<1%
31	Kevin Maik Jablonka, Daniele Ongari, Seyed Mohamad Moosavi, Beren... Crossref	<1%
32	Liverpool John Moores University on 2020-03-03 Submitted works	<1%

turnitin

Similarity Report ID: oid:27535:36266412

33	University College London on 2021-06-25 Submitted works	<1%
34	arxiv.org Internet	<1%
35	biblio.ugent.be Internet	<1%
36	david.ncifcrf.gov Internet	<1%
37	de.slideshare.net Internet	<1%
38	etflin.com Internet	<1%
39	mindmajix.com Internet	<1%
40	molinstincts.com Internet	<1%
41	researchgate.net Internet	<1%
42	"Science of Ashwagandha: Preventive and Therapeutic Potentials", Spr... Crossref	<1%
43	Cho, June-Haeng, Mi-Young Lee, Irshad Ahmed Baig, Na-Reum Ha, Jou... Crossref	<1%
44	Jochen Sieg, Florian Flachsenberg, Matthias Rarey. "In Need of Bias Co... Crossref	<1%

45	Kevser Kübra Kırboğa, Sumra Abbasi, Ecir Uğur Küçüksille. "Explainabil... Crossref	<1%
46	Midlands State University on 2018-06-06 Submitted works	<1%
47	Saad Alghamdi, Yasser Hussein Issa Mohammed, ahmed Hassen Shnt... Crossref posted content	<1%
48	Sanna-Maria Hede, Inga Nazarenko, Monica Nistér, Mikael S. Lindströ... Crossref	<1%
49	Universiti Teknologi Malaysia on 2017-03-10 Submitted works	<1%
50	University College London on 2022-09-22 Submitted works	<1%
51	University of Bath on 2017-01-06 Submitted works	<1%
52	University of Edinburgh on 2020-10-24 Submitted works	<1%
53	University of Greenwich on 2022-08-30 Submitted works	<1%
54	Vels University on 2019-01-21 Submitted works	<1%
55	d.lib.msu.edu Internet	<1%
56	dokumen.pub Internet	<1%

turnitin

Similarity Report ID: oid:27535:36266412

57	mts.intechopen.com Internet	<1%
58	jcheminf.com Internet	<1%
59	researchsquare.com Internet	<1%
60	thno.org Internet	<1%
61	Ipsita Mohanram, Jyotsna Meshram. "Design, synthesis, and evaluatio... Crossref	<1%
62	Patrick K. Jjemba, Brian K. Kinkle, Jodi R. Shann. "In-situ enumeration ... Crossref	<1%
63	Sarah Naomi Bolz, Melissa F. Adasme, Michael Schroeder. "Toward an ... Crossref	<1%
64	University of Derby on 2021-08-10 Submitted works	<1%
65	University of Sydney on 2022-01-04 Submitted works	<1%
66	University of Westminster on 2018-04-30 Submitted works	<1%

Your paper, ID:ICSTSN 194, has been ACCEPTED

1 message

ICSTSN 2023 <icstsn2023@ifet.ac.in> Thu, 23 Mar 2023 at 10:02 pm
To: nidaefalak380@gmail.com, aahraharsh@gmail.com, asmitadas1710@dce.ac.in

Dear Author,

Congratulations!!!

The review and selection process for your paper **ID ICSTSN-194** entitled " **Diospyros kaki's phytochemical mediated inhibition of GBM by targeting PDK-1: An in-silico docking and machine learning model.**" has been completed. Based on the recommendations from the reviewers assigned for your paper, I am pleased to inform you that your paper has been **ACCEPTED by the Technical Program Committee (TPC) for ORAL PRESENTATION** which is organized by IFET College of Engineering, Villupuram, Tamil Nadu, India during 21st - 22nd, April 2023. I am also glad to inform you that the proceedings of ICSTSN 2023 will be submitted for inclusion in IEEE Xplore.

Note : Conference will be held in both OFFLINE and ONLINE MODE.

Registration

-

You are further requested to do the following

- You are requested to kindly **register** at the earliest (after paying the Conference Registration Fees) using the Registration Link.
https://forms.gle/Xxe8VAm3YC1hoepT8

- **Registration will be closed on 27th March 2023.**

- IEEE members can avail the membership benefits on registration fees. Please attach the scanned copy of your IEEE membership card in the Google form.

Final submission Checklist

-

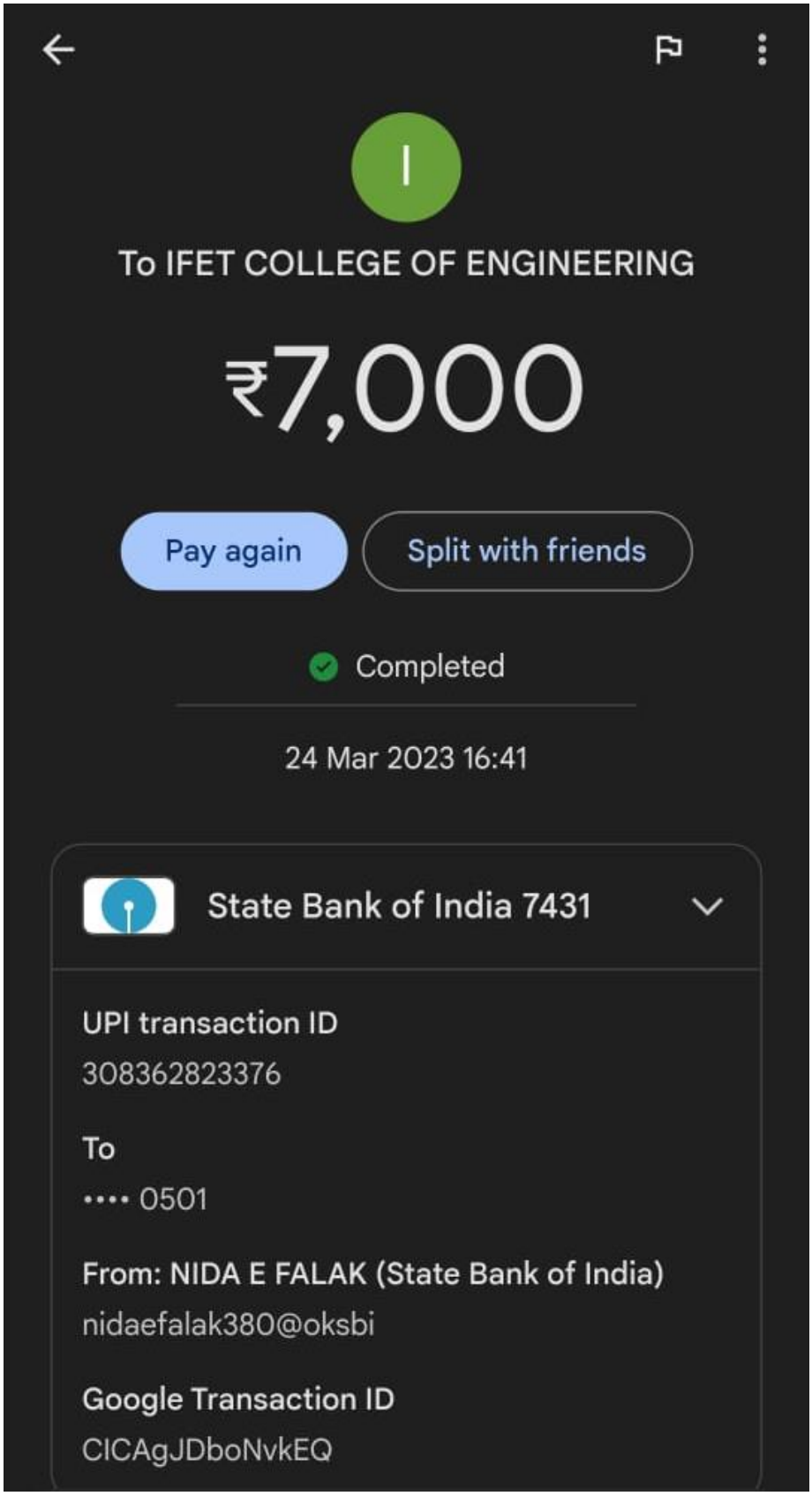
The following documents have to be submitted along with the camera-ready paper on or before 27.03.2023.

1. Camera ready paper in IEEE double column format (in Microsoft office word file) should be uploaded in the CMT portal.
2. Filled in Google form.
3. Proof of registration fee paid.

(The final paper should not exceed 6 pages, IEEE Xplore does not support pages more than 6. Rs.500 will be charged for every extra page)

With Regards,

Conference Coordinator
ICSTSN-2023


To IFET COLLEGE OF ENGINEERING
₹7,000
Pay again
Split with friends
Completed
24 Mar 2023 16:41
State Bank of India 7431
UPI transaction ID
308362823376
To
•••• 0501
From: NIDA E FALAK (State Bank of India)
nidaefalak380@oksbi
Google Transaction ID
CICAgJDboNvkEQ

IFET COLLEGE OF ENGINEERING

AN AUTONOMOUS INSTITUTION

Approved by AICTE, Permanently Affiliated to Anna University,

Recognized under section 2(f) & 12(B) of UGC Act- 1956. Recognized Research Centre of Anna University (ECE)

IFET Nagar, Gangarampalayam, Villupuram - 605 108. Website: www.ifet.ac.in

IEEE MADRAS SECTION

BEST PAPER AWARD

PRESENTED TO

MR. HARSH AAHRA

PG STUDENT, DELHI TECHNOLOGICAL UNIVERSITY

for the paper entitled

DIOSPYROS KAKI'S PHYTOCHEMICAL MEDIATED INHIBITION OF GBM BY INHIBITING PDK-1: AN IN-SILICO AND MACHINE LEARNING MODEL

in

Second IEEE International Conference on *"Smart Technologies and Systems for Next Generation Computing(ICSTSN 2023)"* held at IFET College of Engineering on 21st and 22nd April 2023

CONFERENCE CHAIR PRINCIPAL CHAIRMAN

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi college of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "In Silico and Machine Learning Analysis Unveil Withania somnifera Phytochemicals' Inhibition of GBM through PDK-1 Targeting", which is submitted by Harsh Aahra, 2K21/MSCBIO/16, Delhi Technological University Delhi, in partial fulfilment of the requirement for the award of the degree of Masters in Science, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi:

Date: 02/06/23

DR. ASMITA DAS
SUPERVISOR
Department of biotechnology
Delhi technological university

02/06/2023

PROF. PRAVIR KUMAR
HEAD OF DEPARTMENT
Department of biotechnology
Delhi technological university

III

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi college of Engineering)

Bawana Road, Delhi-110042

DECLARATION

I, Harsh Aahra, 2K21/MSCBIO/16 of MSc. Biotechnology, hereby declare that the project Dissertation titled "**In Silico and Machine Learning Analysis Unveil Withania somnifera Phytochemicals' Inhibition of GBM through PDK-1 Targeting**" which is submitted by me to the Department of Biotechnology, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Science Biotechnology and submitted to the Department of Biotechnology, Delhi Technological University, Delhi, is an authentic record of my own carried out during the period from January -May 2023 under the supervision of Dr. Asmita Das.

The matter presented in this report has not been submitted by me for the award for any other degree of this or any other institute/University. The work has been accepted in SCI/SCI expanded / SSCI/Scopus Indexed Journal OR peer reviewed Scopus Index Conference with the following details :

Title of the Paper : Diospyros kaki's phytochemical mediated inhibition of GBM by targeting PDK-1: An in-silico docking and machine learning model

Authors Name : Harsh Aahra, Nida E Falak, Asmita Das

Name of Conference : Second IEEE International Conference on "Smart Technologies and System for Next Generation Computing (ICSTSN 2023)

Conference Date and Venue : 21st-22nd April 2023 IFET College of Engineering, Gangarampalaiyam, Tamil Nadu, India

Registration: Done

Status of Paper : Under proceeding

Date of Paper Communication: 20th December 2022

Date of Paper Acceptance: 23 March 2023

Date of Paper Publication:

Harsh Aahra

Harsh Aahra

Date: 02/06/23