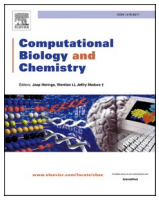
Computational Biology and Chemistry 102 (2023) 107799

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac



OSADHI – An online structural and analytics based database for herbs of India

Kikrusenuo Kiewhuo [a,b], Dipshikha Gogoi [a,b], Hridoy Jyoti Mahanta [a,b], Ravindra K. Rawal [a,b], Debabrata Das [a,b], Vaikundamani S [a], Esther Jamir [a], G. Narahari Sastry [a,b,*]

[a] Advanced Computation and Data Sciences Division, CSIR-North East Institute of Science and Technology, Jorhat 785006, Assam, India
[b] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, Uttar Pradesh, India

ARTICLE INFO

Keywords:
Medicinal plants
Phytochemicals
ADMET
Traditional Knowledge
Cheminformatics

ABSTRACT

The current study aims to develop a PAN India database of medicinal plants along with their phytochemicals and geographical availability. The database consists of 6959 unique medicinal plants belonging to 348 families which are available across 28 states and 8 union territories of India. The database sources the information on four different sections – traditional knowledge, geographical indications, phytochemicals, and chemoinformatics. The traditional knowledge reports the plant taxonomy with their vernacular names. A total of 27,440 unique phytochemicals associated with these plants were curated from various sources in this study. However, due to the non-availability of general information like IUPAC names, InChI key, etc. from reliable sources, only 22,314 phytochemicals have been currently reported in the database. Various analyses have been performed for the phytochemicals which include analysis of physicochemical and ADMET properties calculated from open-source web servers using in-house python scripts. The phytochemical data set has also been classified based on the class, superclass, and pathways respectively using NPClassifier, a deep learning framework. Additionally, the antiviral potency of the phytochemicals was also predicted using two machine learning models – Random Forest and XGBoost. The database aims to provide accurate and exhaustive data of the traditional practice of medicinal plants in India in a single platform integrating and analyzing the rich customary practices and facilitating the development and identification of plant-based therapeutics for a variety of diseases. The database can be accessed at https://neist.res.in/osadhi/.

1. Introduction

The Indian sub-continent comprises of over 45,000 plant species with two out of the four mega diversity areas of the world (Singh et al., 2019; Mahomoodally et al., 2019). India's medicinal plant wealth plays a crucial role in maintaining the health of the population, especially the ethnic groups. The country has used numerous plants and herbs for the treatment of many different ailments and continues to uphold the practice in the form of Ayurveda, Homeopathy and Naturopathy, Siddha, Unani, and Yoga for centuries. Plant-based medicine significant impact on the healthcare sector where about 80 % of the world population depend on them (Pešić and Stanković, 2015; Kumar, 2020). The traditional system of medicine has a huge impact on the public healthcare and it is intricately intertwined with festivals and religions (Prabhu et al., 2014; Thomford et al., 2018). Documenting the traditional uses of medicinal plants can be either systemic or non-systemic. The systemic approach refers to well-documented literature such as Ayurveda, Sidha etc. (Mukherjee et al., 2017; Singh et al., 2020). The transition from Atharva Veda's religious therapies to the classical Ayurvedic medicines has taken place in a systematic approach (Krishna, 2022). Whereas, for non-systemic refers to the practices used by the ethnic communities of a region and are orally documented. Proper documentation of the non-systemic approaches of traditional practices is a primary requirement as it may get lost over time.

Medicinal plants are the principal sources of various phytochemicals which are often used as a component of natural product-based drugs (Ugboko et al., 2020; Ram et al., 2022). Some of the natural products from medicinal plants commonly used in the production of drugs are strychnine, cocaine, nicotine, papaverine, quinine, digitoxin etc. (Aslam and Ahmad, 2016; Harvey et al., 2015). There are about 100 plant-based compounds currently under various phases of clinical trials and the search for new drugs for various life-threatening diseases continues to

* Corresponding author at: Advanced Computation and Data Sciences Division, CSIR-North East Institute of Science and Technology, Jorhat 785006, Assam, India.
E-mail addresses: gnsastry@gmail.com, gnsastry@neist.res.in (G.N. Sastry).

https://doi.org/10.1016/j.compbiolchem.2022.107799
Received 27 September 2022; Received in revised form 27 November 2022; Accepted 28 November 2022
Available online 1 December 2022

Table 1
A feature-wise comparison of existing Indian medicinal plants databases with OSADHI.

Databases	Medicinal herb Database	NeMedPlant	MEDDB	Phytochemica	MED-PDB	AromaDb	UMPDB	IMPDB	NEIMPDB	IMPPAT	OSADHI
Geographic Coverage	Whole India	Manipur & Nagaland	Whole India	Whole India	Whole India	Whole India	Uttarakhand	Whole India	North East India	Whole India	Whole India
No of plants	1915	56	110	561	147	165	1127	113	561	4010	6959
No of Phytochemicals	–	–	–	963	369	–	1322	1829	9410	17,967	27,440
Year of launch	–	2012	2012	2015	2017	2018	2018	2022	2022	2022	2022
Taxonomic classification	√	√	√	√	√	√	√	√	√	√	√
Vernacular name	√	X	X	X	X	X	X	X	√	X	√
Synonyms	X	X	X	X	X	X	X	X	X	√	√
Therapeutic uses	√	√	√	X	√	√	√	√	√	√	√
Geographical location indication	X	√	√	X	X	X	√	√	√	X	√
Phytochemical constituents	X	X	X	√	√	√	√	√	√	√	√
Physicochemical properties	X	X	X	√	X	√	X	X	√	√	√
ADMET	X	X	X	√	X	√	X	X	√	√	√
2D/3D structures	X	X	X	√	X	√	X	X	√	√	√
Web-link	https://bsi.gov.in/page/en/medicinal-plant-database	http://bif.uohyd.ac.in/nemedplant/index.php	NA	home.iitj.ac.in/~bagler/webservers/Phytochemica	http://genomeinformatics.dtu.ac.in/medicinalplant/	http://bioinfo.cimap.res.in/aromadb/	NA	https://impdb.org.in/index.php	https://neist.res.in/neimpdb/	https://cb.imsc.res.in/imppat/home	https://neist.res.in/osadhi

grow (Harvey, 2008). Natural products can be a great alternative in the development of effective drugs (Scotti et al., 2021; Xu et al., 2020). Therefore, the proper documentation of these practices will help ensure the preservation and conservation of traditional knowledge.

Medicinal plant database functions as an 'information bank' allowing quick recovery of the information provided by the developers (Vinayak et al., 2010; Manhã et al., 2008). The data contained in these databases are applicable for the analysis of various parameters and are not confined to drug discovery and conservation of the traditional knowledge (Skoczen and Bussmann, 2006). Despite the popular use of these practices, the exact estimate of the medicinal plants reported is underrepresented (Pouliot, 2011; Fitzgerald et al., 2020). In accordance with the documentation of the medicinal plants, various databases have been developed by major stakeholders with an aim to spread the traditional knowledge and help in their conservation. Despite the development of several medicinal plants database, very few remain functional due to non-maintenance and updates of the digital data (Ningthoujam et al., 2012).

The current study focuses on the documentation and digitalization of medicinal plants available in all the states and union territories of India. In earlier days, "*herb*" was applied only to non-woody plants, but nowadays, it refers to any part of the plant like fruit, root, seed, stem, leaf, stigma, etc (https://www.nhp.gov.in/). The Indian meaning of the OSADHI (or OSHADHI or AUSHADHI) means "*medicinal herbs*", however in this study we have used it as an acronym for "An **O**nline **S**tructural and **A**nalytics based **D**atabase for **H**erbs of **I**ndia". Considering the modern meaning of "*herb*", the current study refers to all the plants which can be used for medicinal purposes. With reference to this, OSADHI serves as an open-access database that provides information like taxonomic classification, parts used, and therapeutic indications of 6959 medicinal plants available across the states and union territories of India. The geographical availability of each medicinal plant has also been indicated covering all the twenty-eight states and eight union territories of India. An important feature of the database is the exhaustive data on the phytochemicals reported on the medicinal plants in various literature and open-access databases. The database currently houses a total of 22,314 unique phytochemicals present in 6959 medicinal plants which have been curated from 1,64,994 phytochemicals. The physicochemical, medicinal chemistry and ADMET properties of 22,314 phytochemicals have been calculated through various open-access servers using in-house python scripts. Apart from these, the phytochemicals were also subjected to various classifications based on pathways, superclass, and class. An important aspect of the database is that it also predicts the antiviral potency of the phytochemicals using machine learning models. Along with all this information, the 2D and 3D structures of the compounds are also available to users and can be accessed using https://neist.res.in/osadhi/. Table 1 demonstrates a feature-wise comparison of OSADHI with databases developed by various research groups across the country.

2. Methodology

2.1. Data collection and curation

Data on medicinal plants, geographical location, and traditional practices were mined from published literature of major journals and reliable existing databases published between 1986 and 2022 (Srivastava et al., 1986; Dey and De, 2012; Parthiban et al., 2016; Meetei et al., 2012; Mary et al., 2012; Pathania et al., 2015; Sargia et al., 2018; A. Kumar et al., 2018; Y. Kumar et al., 2018; Saldívar-González et al., 2019; Sinha et al., 2022; Kiewhuo et al., 2022; Vivek-Ananth et al., 2022). Each source was queried using the state and union territory name as search term from Google scholar (https://scholar.google.com), PubMed (https://pubmed.ncbi.nlm.nih.gov), National Medicinal Plants Board (https://nmpb.nic.in/), AYUSH (https://www.ayush.gov.in/), and Researchgate (https://www.researchgate.net/). The initial information

Fig. 1. Schematic workflow used in the development of **O**nline **S**tructural and **A**nalytics based **D**atabase for **H**erbs of **I**ndia (OSADHI).

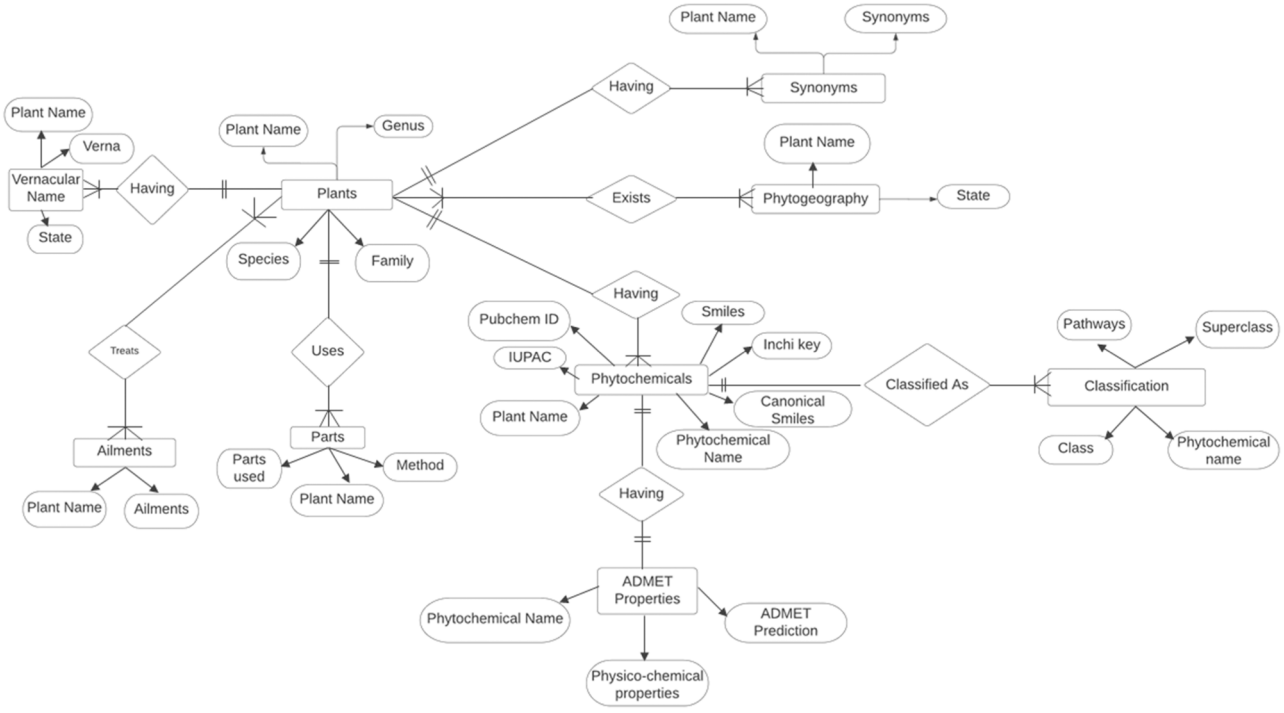


Fig. 2. The entity relationship (ER) diagram of OSADHI.

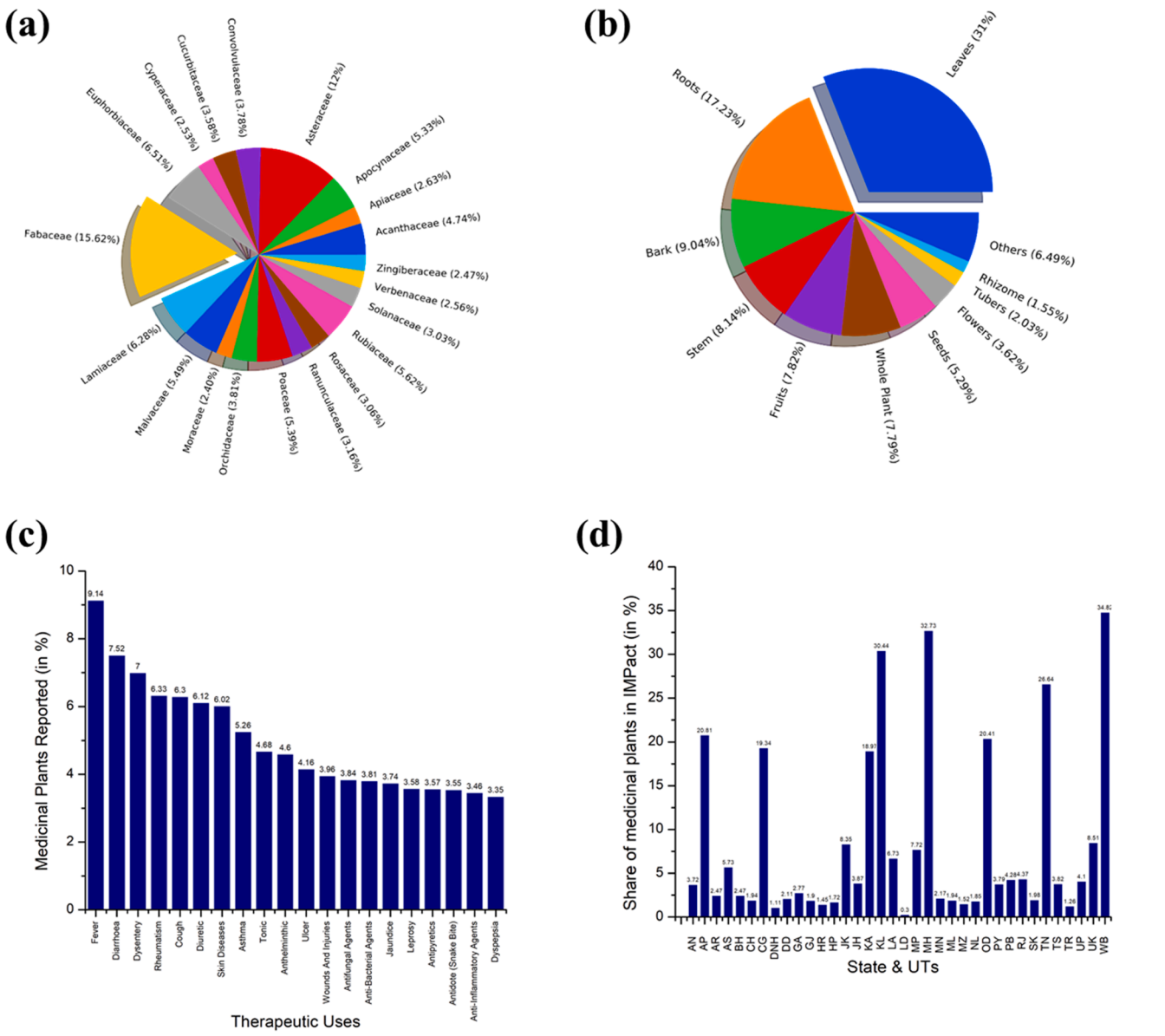


Fig. 3. Details of the distribution of the medicinal plants (a) Families (b) Parts used (c) Therapeutic uses and (d) Geographical availability.

on the medicinal plants like taxonomy information, vernacular names, and geographical availability, were manually curated with minimal errors and then were clustered based on their state-wise availability. The data on the phytochemicals associated with each plant was obtained from IMPPAT (https://cb.imsc.res.in/imppat/home), Dr. Duke's Phytochemical Database (https://phytochem.nal.usda.gov/phytochem/search), Phytochemica (http://faculty.iiitd.ac.in/~bagler/webservers/Phytochemica/), and literature mining. The medicinal plants and phytochemicals data also under went multiple redundancy checks using in-house python scripts and finally, the unique data (6959 medicinal plants and 27,440 phytochemicals) were identified and subjected to further analysis.

Upon the identification of the list of unique phytochemicals, detailed information such as the IUPAC name, SMILES, InChI Key and PubChem ID was retrieved from PubChem (https://pubchem.ncbi.nlm.nih.gov/) using PubChem Identifier Exchange Service (https://pubchem.ncbi.nlm.nih.gov/idexchange/idexchange.cgi) (Hur and Wild, 2008). This is an open-source server that converts the input into the selected output operation such as SMILES, InChI key, synonyms, compound ID etc. Other sources like SpectraBase (https://spectrabase.com/), FooDB (https://foodb.ca/), ChEMBL (https://www.ebi.ac.uk/chembl/), and GreenmolBD (https://www.greenmolbd.gov.bd/) were also used for the compounds whose details were not directly available in PubChem. A schematic representation of the workflow is shown below in Fig. 1.

2.2. Features of the database

The various features of the medicinal plants available in the database include taxonomic classification, families, vernacular name, traditional practices (parts and method used, ailments treated), state and union territory of availability, synonyms, and phytochemical details of the medicinal plant. For the phytochemicals, the retrieved data includes the compound name (trivial and IUPAC name), PubChem IDs, canonical/isomeric SMILES, InChI key, physicochemical properties, ADMET properties, antiviral property, medicinal chemistry, and 2D/3D structure.

2.2.1. Phytochemical classification

The classification of the phytochemicals into appropriate classes, super-class, and pathways was performed using NP Classifier (https://

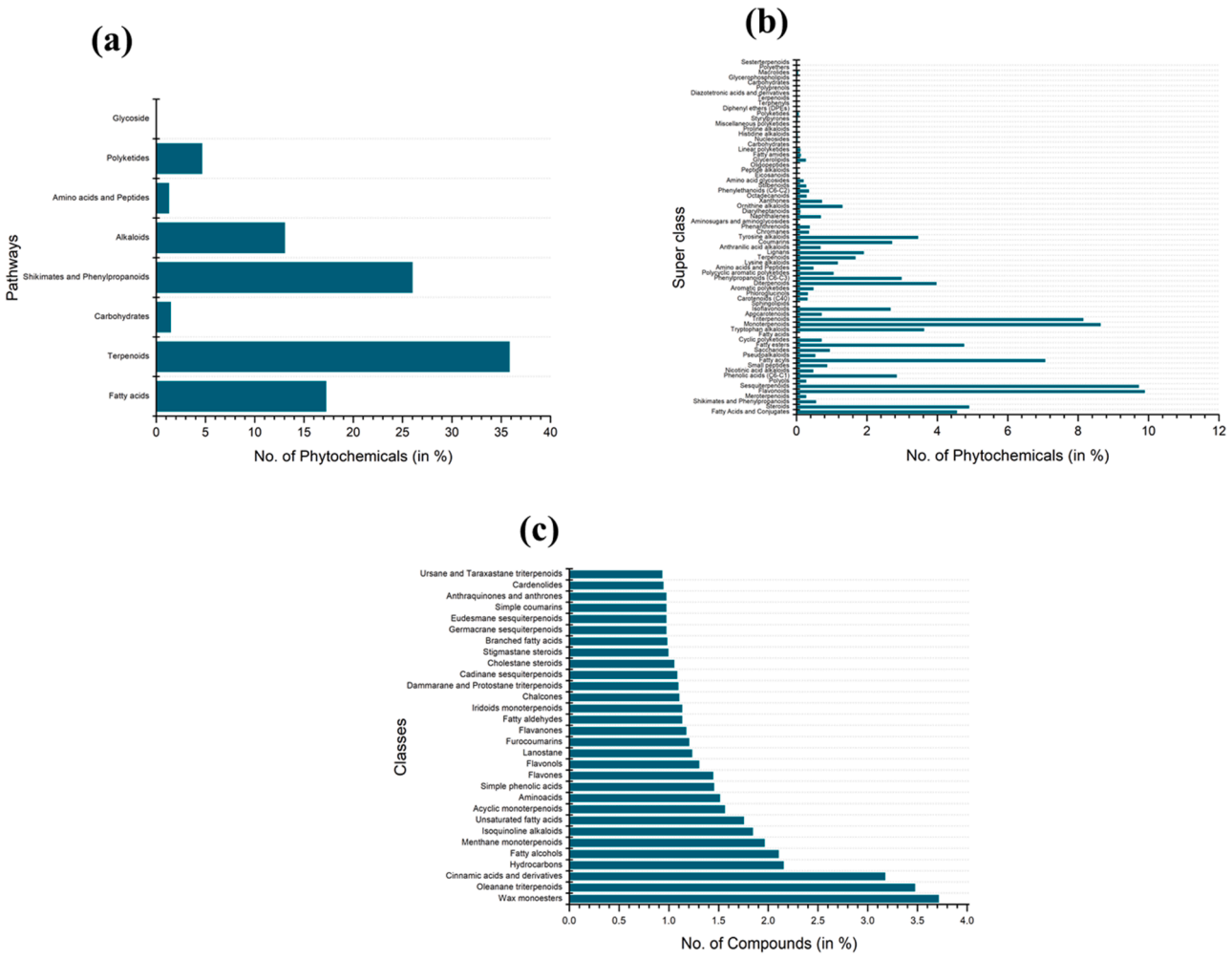


Fig. 4. Distribution of phytochemical compounds based on (a) Pathways (b) Superclass and (c) Class calculated using NP Classifier tool.

npclassifier.ucsd.edu) [Kim et al. (2021)]. NP Classifier is a web server that was designed for the structural classification of natural products using deep learning. The tool provides users with information on the metabolism, chemical properties, and structural details of natural products. NP classifier can classify a compound into 7 different pathways such as alkaloids, terpenoids, fatty acids, polyketides, shikimates−phenylpropanoids, amino acids/peptides, and carbohydrates, 70 super-classes and 653 classes based on their semantic knowledge. NP Classifier was accessed using their publicly available APIs and the classes were extracted through an in-house python script with canonical SMILES as inputs.

2.2.2. *Physicochemical properties*

The descriptor calculation was performed using Modred (Moriwaki et al., 2018) and RDKit [https://www.rdkit.org/docs/index.html] packages in python environment. Mordred generates around 1600 and RDKit generates 208 descriptors respectively. Some of the physiochemical properties which have been reported in this study include molecular weights, solubility, partition coefficient, various atom counts, various bond counts, ring counts, and many more which depict the general as well as structural analytics of the phytochemicals.

2.2.3. *ADMET properties*

In silico calculation of ADMET properties is widely applied in drug discovery (Kar and Leszczynski, 2020). It determines the efficiency and efficacy of a drug molecule (Gola et al. (2006); Fan and Lannoy, 2014). The ADMET properties of 22,314 phytochemicals were calculated using ADMETlab 2.0 (https://admetmesh.scbdd.com/) (Xiong et al., 2021). It is a redesigned version of ADMETlab webserver which calculates the pharmacokinetic and toxicity properties of the compounds. The present version calculates a total of 88 different properties under absorption, distribution, metabolism, excretion, and toxicity. For the current study, following properties have been considered and documented in the database - Caco-2 Permeability, MDCK Permeability, HIA for absorption; Pgp-inhibitor, Pgp-substrate, PPB, VD, BBB Penetration, and Fu for distribution; CYP 1A2 inhibitor, CYP 1A2 substrate, CYP 2C19 inhibitor, CYP 2C19 substrate, CYP 2C9 inhibitor, CYP 2C9 substrate, CYP 2D6 inhibitor, CYP 2D6 substrate, CYP 3A4 inhibitor, and CYP 3A4 substrate for metabolism; CL and t1/2 for excretion; and AMES Toxicity, Rat Oral Acute Toxicity, Carcinogenicity, Respiratory Toxicity, NR-Aromatase, hERG-blockers, Eye Irritation, NR-PRAR-gamma, Skin sensitization, and FAF-Drugs4 Rule for the toxicity respectively.

2.2.4. *Medicinal chemistry*

Medicinal chemistry aids in the identification of new leads, and the optimization of clinical candidates (Scotti and Scotti, 2019). Its properties help determine if a compound/substance fulfils a drug profile. The medicinal properties considered for this study include QED – a measure of drug-likeness, SAscore - estimation of ease of synthesis, and NPscore - natural product-likeness were also calculated. The drug-likeliness of the phytochemicals was also calculated using Lipinski Rule, Golden Triangle Rule, Pfizer Rule, and GSK rule (Chen et al., 2020). These properties were also calculated using ADMETlab 2.0 sever.

Table 2
List of 20 highest occurring phytochemicals among the 6959 medicinal plants of India as identified in OSADHI database.

Sl. No.	Phytochemical	IUPAC	Frequency (in %)
1.	β-Sitosterol	*(3 S,8 S,9 S,10 R,13 R,14 S,17 R)*− 17-[*(2 R,5 R)*− 5-Ethyl-6-methylheptan-2-yl]− 10,13-dimethyl-2,3,4,7,8,9,11,12,14,15,16,17-dodecahydro-1*H*-cyclopenta[*a*]phenanthren-3-ol	23.64
2.	α-Pinene	2,6,6-Trimethylbicyclo[*3.1.1*]hept-2-ene	14.37
3.	Limonene	1-Methyl-4-prop-1-en-2-ylcyclohexene	13.49
4.	beta-Caryophyllene	(1 *R*,4*E*,9 *S*)− 4,11,11-Trimethyl-8-methylidenebicyclo[*7.2.0*]undec-4-ene	13.33
5.	Linalool	3,7-Dimethylocta-1,6-dien-3-ol	12.78
6.	β -Pinene	6,6-Dimethyl-2-methylidenebicyclo [*3.1.1*]heptane	12.11
7.	Myrcene	7-Methyl-3-methylideneocta-1,6-diene	11.69
8.	α-Terpineol	2-(4-Methylcyclohex-3-en-1-yl)propan-2-ol	11.38
9.	*p*-Cymene	1-Methyl-4-propan-2-ylbenzene	11.26
10.	Palmitic acid	Hexadecanoic acid	11.17
11.	4-Carvomenthenol	4-Methyl-1-propan-2-ylcyclohex-3-en-1-ol	10.61
12.	Humulene	(1*E*,4*E*,8*E*)− 2,6,6,9-Tetramethylcycloundeca-1,4,8-triene	10.58
13.	Camphene	2,2-Dimethyl-3-methylidenebicyclo [2.2.1]heptane	10.15
14.	Caryophyllene oxide	(1 *R*,4 *R*,6 *R*,10 *S*)− 4,12,12-Trimethyl-9-methylidene-5-oxatricyclo[*8.2.0.04,6*] dodecane	9.84
15.	Eucalyptol	1,3,3-Trimethyl-2-oxabicyclo[*2.2.2*] octane	9.82
16.	Quercetin	2-(3,4-Dihydroxyphenyl)− 3,5,7-trihydroxychromen-4-one	9.44
17.	γ-Terpinene	1-Methyl-4-propan-2-ylcyclohexa-1,4-diene	9.16
18.	(+)-delta-Cadinene	(1 *S*,8*aR*)− 4,7-Dimethyl-1-propan-2-yl-1,2,3,5,6,8*a*-hexahydronaphthalene	9.15
19.	Sabinene	4-Methylidene-1-propan-2-ylbicyclo [*3.1.0*]hexane	9.06
20.	Terpinolene	1-Methyl-4-propan-2-ylidenecyclohexene	9.00

2.2.5. Prediction of antiviral compounds

One of the unique features of the database includes the antiviral potency of the phytochemicals. This was calculated using two machine learning models - Random Forest and XGBoost from John et al. (2022). Different hyperparameters were used for the optimization of the models (Random Forest, XGBoost, Support Vector Machine, KNN, Decision Tree, MLP Classifier, and Logistic Regression) considered in this study. However, it was found that the accuracy and reliability of the predictions from the two models i.e. Random Forest and XGBoost were higher compared to other models. Additionally, these two models have also given reliable predictions when tested over the external validation set (known antivirals). Hence, these two models (Random forest and XGBoost) were used to screen the compounds in OSADHI to predict their antiviral potency. The descriptors calculated for the phytochemicals were preprocessed and the selected descriptors were considered as input to the predictive models. As both the models performed best in antiviral prediction (John et al., 2022), the consensus outputs from both the models were considered to arrive at the prediction as well as the prediction probability score. A large set was found to give the same prediction with both the models and so the highest probability score was also considered. For those in which the predictions were complementing, the one with the highest probability scores was considered the final prediction.

2.2.6. Database design

The relational database MySQL is used in the development of OSADHI to facilitate organized and effective content management. PHP (Hypertext Preprocessor) is used to construct a user-friendly interface, which consists of dynamic web pages for managing and visualizing data. Visualizing tool like Jmol, a Java-based applet tool has been incorporated into the chemoinformatics module for generating the structures dynamically. The front-end of OSADHI has been developed using four core technologies such as Bootstrap, CSS (Cascading Style Sheets), HTML (Hypertext Markup Language), and JavaScript. Bootstrap is a free open source front end development framework that is used to design responsive websites in an easier way. It includes HTML and CSS templates for various HTML features like tables, image slides, models, navigations, navbar etc. CSS (Cascading Style Sheets) is a style sheet language used for a better presentation of HTML elements in a webpage and user interface. It is written in different markup languages like HTML, XML etc. HTML (Hyper Text Markup Language) is a makeup language used for the development of webpages for display in web browsers. JavaScript is used for making dynamic web pages which uses client-side script that runs in the browser for real-time interaction with the user.

A total of nine tables were built and included in the current database, which serves as a repository for data on medicinal plants, their therapeutic uses on family, scientific names, vernacular names, availability in various regions, plant part used for therapeutic purposes, method of use, phytochemicals and their physicochemical, ADMET properties, and classifications. To avoid any sort of anomalies, each table has been normalized to the third normal form (3NF). To preserve a one-to-one relationship for the purpose of making the search queries easier and to obtain the query results, primary indexing was implemented. This front-end gives users the option to search for plants by their scientific name, therapeutics, family, phytochemicals, and natural product classes. Fig. 2 shows the entity-relationship (ER) diagram of OSADHI.

3. Results and discussions

3.1. Database structure

The database structure has been segmented into four features - Traditional knowledge, geographical location, phytochemicals, and chemoinformatics.

3.1.1. Traditional knowledge

The traditional knowledge feature of the database comprises of information on the taxonomic classification. This feature also includes information on traditional practices of medicinal plants and the various ailments treated using them. The taxonomic classification of 6959 medicinal plants has been listed in this section. The geographical availability of these plants helped to identify the vernacular name of around 6860 medicinal plants and has also been reported in the current study. This feature will be helpful to find out the availability of a particular medicinal plant by the local people of a specific area. The database also records a total of 343 unique families belonging to 6959 medicinal plants. The family Fabaceae had the maximum number of medicinal plants (467), followed by Asteraceae (364), and Lamiaceae (337) (Fig. 3a). The parts used, and methods used for the treatment of various ailments have also been documented in the database. The use of a total of 40 different combinations of plant parts has been recorded of which leaves are most commonly used (31 %). 17.23 % of medicinal plants used roots for the treatment of ailments. The corm of *Gloriosa superba* and *Colchicum luteum* were some of the least recorded parts used (Fig. 3b). The ailments cured using these medicinal plants by the ethnic communities have also been documented. A combination of 2477 different ailments have been reported, of which fever was the most commonly treated ailment (1356). This was followed by diarrhea (1158) and dysentery (1014) respectively (Fig. 3c). The top twenty ailments treated using medicinal herbs are fever, diarrhea, dysentery, skin diseases, cough, asthma, rheumatism, diuretic, anti-pyretic, astringent, ulcer, tonic, anthelminthic, headache, jaundice, appetite stimulant,

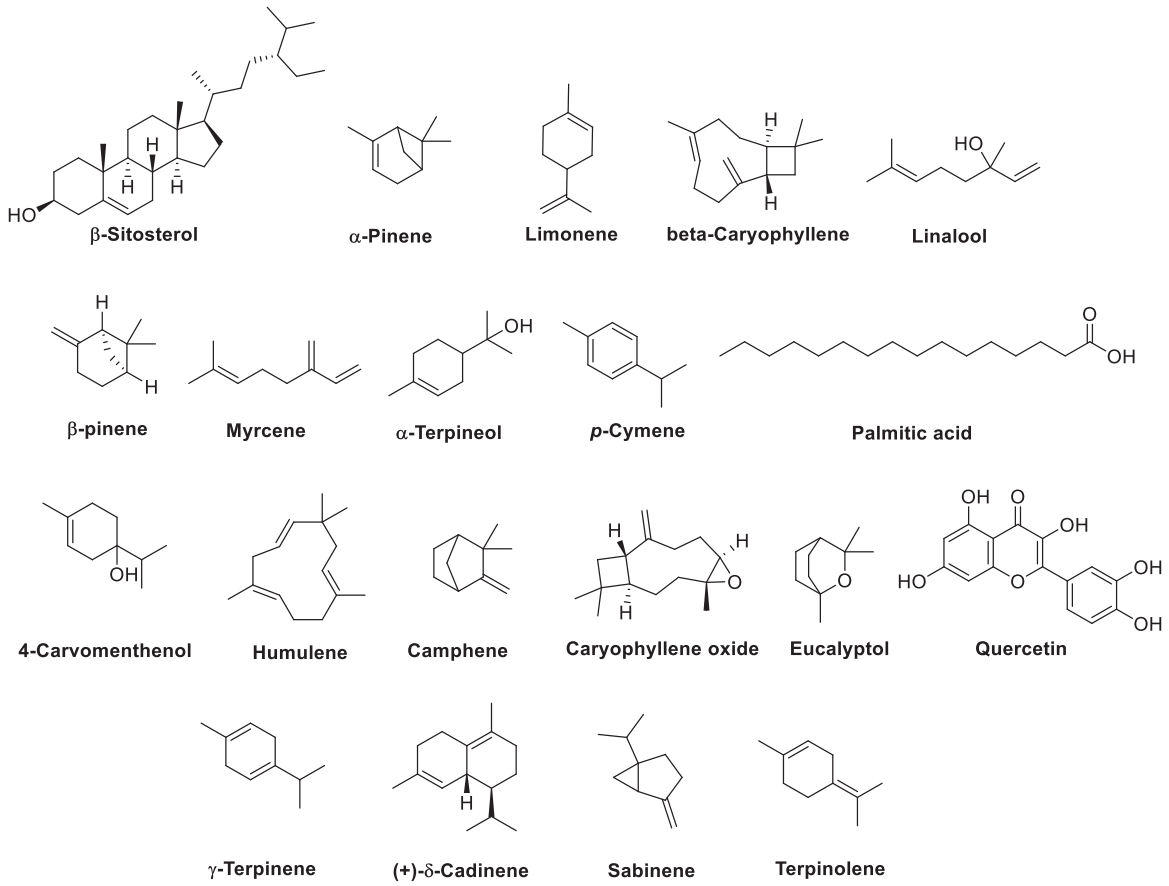


Fig. 5. 2D-Chemical structures of top 20 most occurring phytochemicals in OSADHI database.

leprosy, analgesics, dyspepsia, and flatulence.

3.1.2. Geographical location

The geographical location of 20,786 medicinal plants distributed across twenty-eight states and eight union territories of India have been reported in this study. The highest number of medicinal plants have been recorded in the state of West Bengal (2423), followed by Maharashtra (2278), and Kerala (2118) (Fig. 3d). Some medicinal plants that are well distributed across the country such as: *Centella asiatica (31), Azadirachta indica (30), Mimosa pudica (29), Abrus precatorius (27), Aegle marmelos (27), Ricinus communis (27), Cassia fistula (27), Achyranthes aspera (26), Asparagus racemosus (26),* and *Mangifera indica (25)* (Shukla and Chakravarty, 2012; Jain et al. (2005); Bansal et al. (2016); Reddy et al. (2010); Savithramma et al. (2015); Sharma et al. (2014)).

3.1.3. Classifications

The classification of the phytochemicals was done using the NP classifier. Out of 22,314 phytochemicals, it was found that a large percentage of them belonged to terpenoids (36 %), which was followed by shikimates and phenylpropanoids (26 %). On the other hand, amino acids, carbohydrates, and glycosides had the least phytochemicals as pathways. Similarly, from the superclasses it was observed that most of the phytochemicals belonged to flavonoids, sesquiterpenoids, monoterpenoids, and triterpenoids. They covered major clusters of the phytochemicals reported in this database. And finally looking at the classes, wax monoesters, oleane tepenoids, cinnamic acids, and derivates were the top classes to which a large portion of the phytochemicals was classified. Ursane, cardenolides and anthraquinones and anthrones were the classes having the least number of phytochemicals. Fig. 4 shows the distribution of the phytochemicals with respect to pathways, superclasses and classes as derived from the NP classifier.

3.1.4. Analysis of physiochemical properties

The current version of the database reports 22,314 unique phytochemicals associated with the 6959 medicinal plants with general information like SMILES, InChI key, PubChem ID, IUPAC name, and various physiochemical properties. Initial analysis based on the frequency of occurrences of phytochemicals shows beta-sitosterol, alpha-pinene, limonene, beta-caryophyllene, linalool, beta-pinene, were most common. Table 2 depicts a list of top 20 most occurring phytochemicals in all the medicinal plants. The chemical structure of these top twenty phytochemicals is separately shown in Fig. 5. As these phytochemicals were from plants, the structures show the presence of OH and H.

Analyzing the physicochemical properties, it was observed that the number of rings, aliphatic and saturated rings in the entire set ranged between 0 and 38, while the number of aromatic rings ranged between 0 and 18. The number of hetero rings ranged the highest between 0 and 191 (Fig. 6). Similarly, the number of carbocycles and heterocycles were also calculated for aliphatic, aromatic and saturated cycles. The number of cycles ranged between 0 and 38 with aromatic heterocycles having the lowest range between 0 and 6 (Fig. 7). 92 % of the phytochemicals were found to have the ideal number of hydrogen bond donors and acceptors. It was also observed that 88 % of the dataset had the optimal range of the number of rotatable bonds, and 91 % had the number of rigid bonds respectively (Fig. 8). About 90 % of the calculated phytochemicals fall under the optimal molecular weight range (100−600) and 84 % had the topological polar surface area (TPSA) optimal range. The logP, logS, and logD values were also calculated and it was observed that 45 %, 57 %, and 35 % of the phytochemicals were in the proper range of logP (0–3 log mol/L), logS (−4 to 0.5 log mol/L), and logD (1–3 log mol/L) respectively (Fig. 9). Further, the number of atoms and counts for nitrogen, carbon, oxygen, and hydrogen were also calculated as shown in Fig. 10. It can be seen that almost all the phytochemicals had

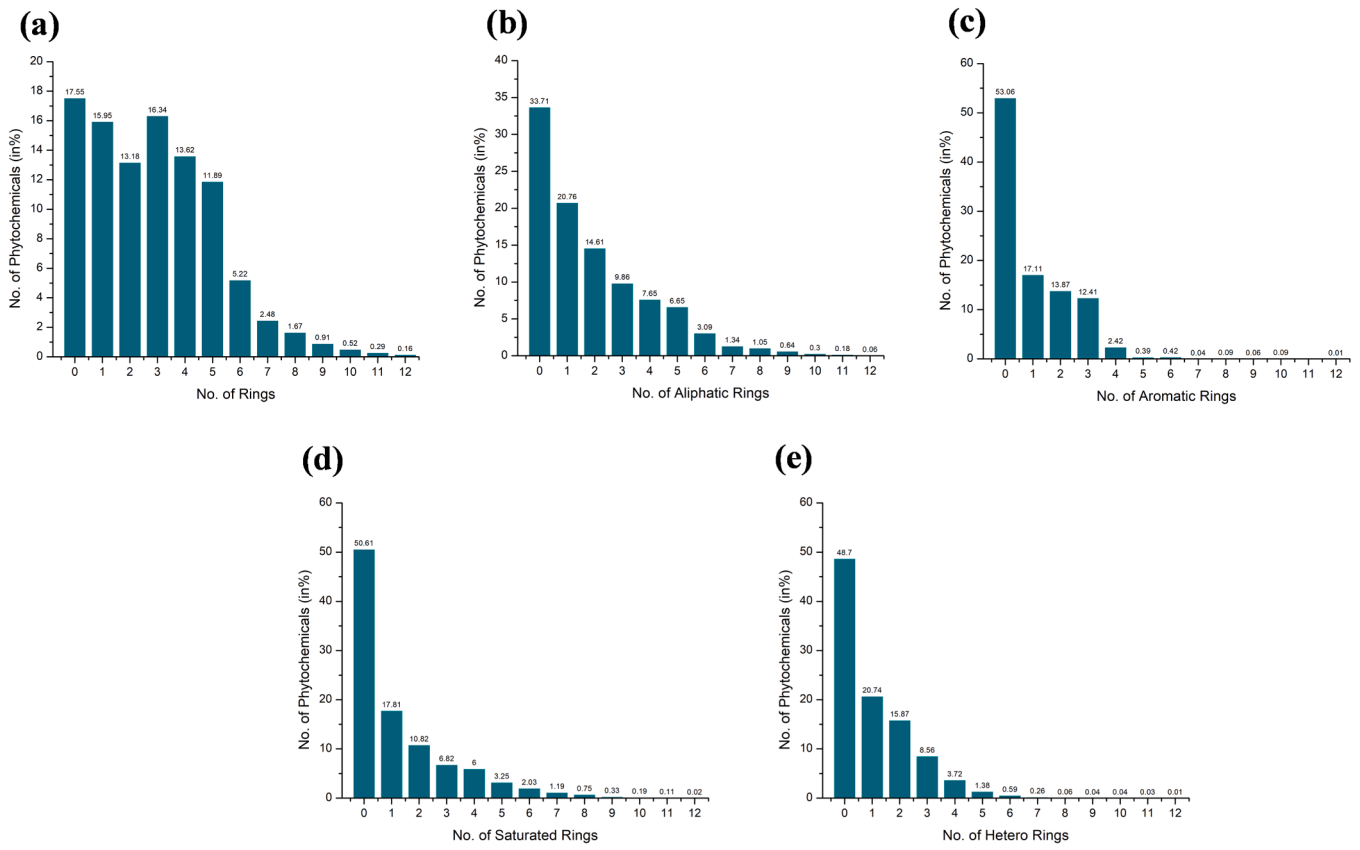


Fig. 6. The presence of different types of rings in the phytochemicals (in %) where (a) 82.45 % of phytochemicals contain rings between 1 and 12 (b) 66.29 % of phytochemicals contain aliphatic rings between 1 and 12 (c) 46.94 % phytochemicals contain aromatic rings between 1 and 4 (b) 49.39 % phytochemicals contain 1–8 saturated rings and (e) 51.3 % phytochemicals contain 1–6 hetero rings, respectively.

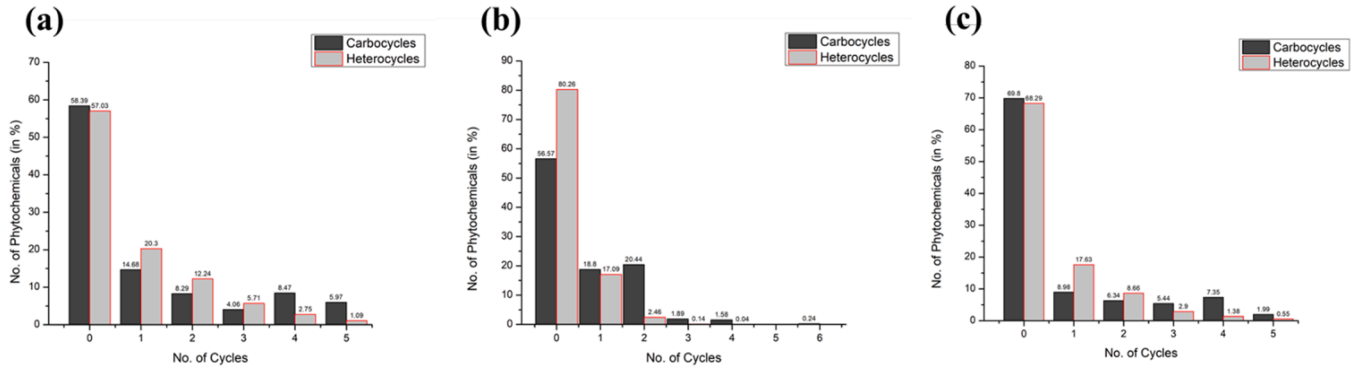


Fig. 7. Representation of the presence of different types of cycles in the phytochemicals (a) Aliphatic (b) Aromatic and (c) Saturated cycles.

Nitrogen atoms (between 0 and 10) and 80 % has Oxygen atoms, the fact is validated because these phytochemicals are available in naturally occurring plants.

3.1.5. Medicinal chemistry

Only 13 % of phytochemical compounds were observed to be attractive compounds based on the QED (Quantitative Estimate of Druglikeness) score which is calculated based on the molecular weight (MW), LogP, topological polar surface area (TPSA), nHBA, nHBD, nRotb, the number of aromatic rings (nAr), and the number of alerts for undesirable functional groups. The synthetic accessibility score (SAscore) was calculated where a low score indicates a compound that is easy to synthesize. About 94 % of the compounds were observed to have scores < 6 which are predicted to be feasible to synthesize. Similarly, the natural product likeness score was also computed to predict the probability of the compound being a natural product. The results exhibited that all the compounds followed the range from -5–5 which indicates all phytochemicals calculated are natural products. The drug-likeliness score of the phytochemicals was also calculated using the following four rules –Lipinski rule, GSK rule, Pfizer rule, and golden triangle rule, respectively. Lipinski rule describes the druggability of a molecule when MW of the molecule is less than 500 Da, logP less than 5, and H acceptor and donor not less than 10 and 5, respectively. Glaxo Smith Kline (GSK) rule accepts the druggability of a molecule when the MW is less than 400 Da and logP value less than 4. Pfizer rule considers the logP value less than 3 and TPSA less than 75, while the golden triangle rule takes MW between 50 and 200 Da and logD value between -2 and 5, respectively. It was observed that 82 % of phytochemicals passed the

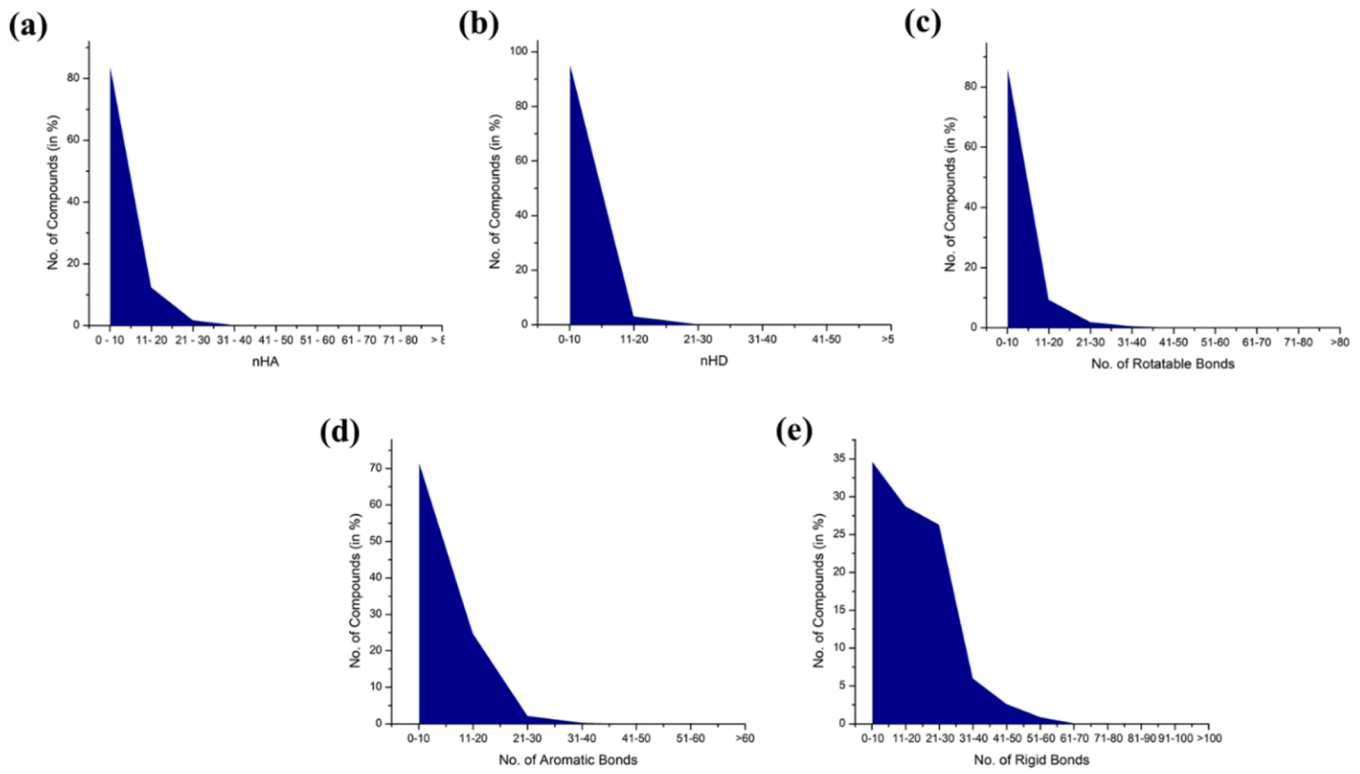


Fig. 8. Calculation of various bonds present in the phytochemicals (a) Number of H-bond acceptor (b) Number of H-bond donor (c) Number of rotatable bonds (d) Number of aromatic bonds, and (e) Number of rigid bonds.

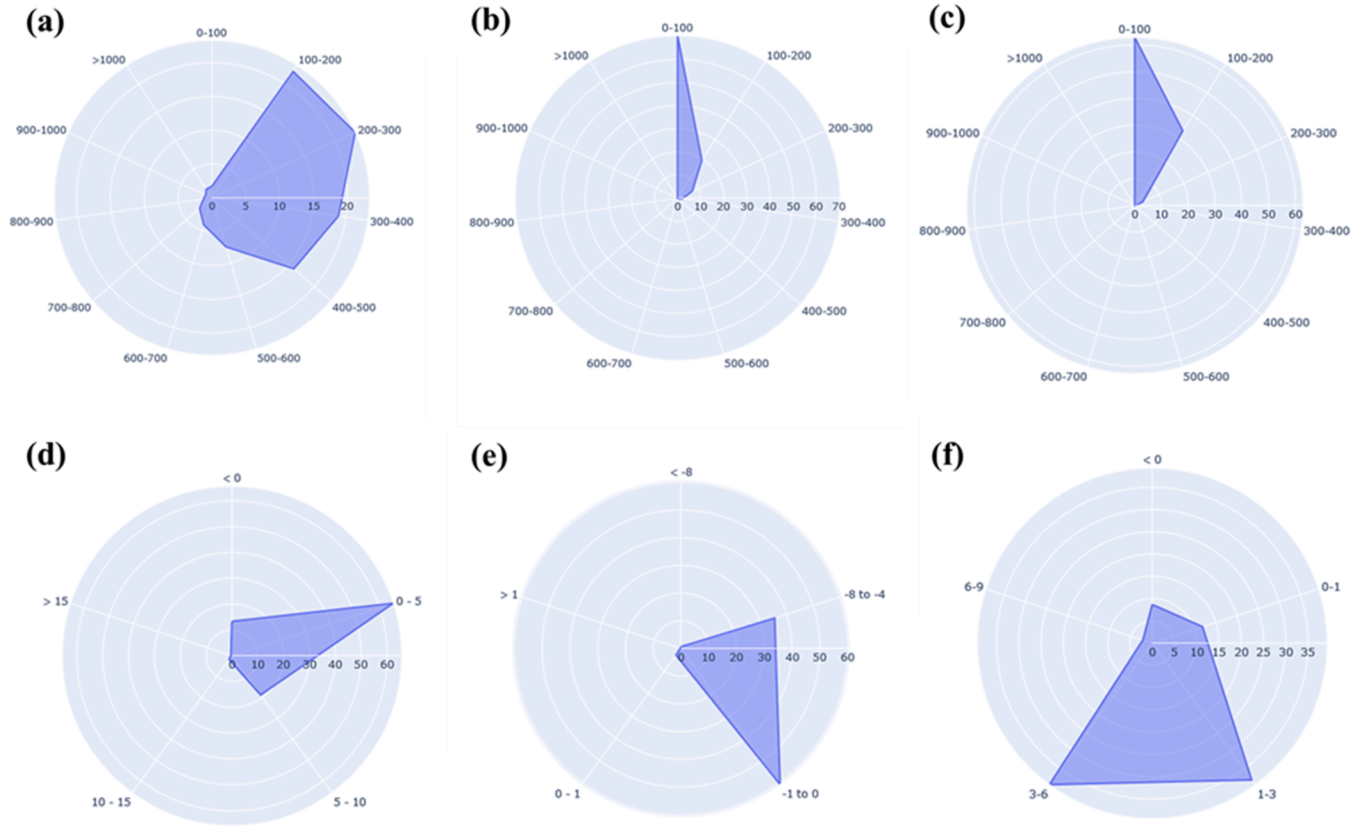


Fig. 9. Distribution and coverage of the phytochemicals depend on (a) Molecular weight (b) TPSA (c) Molar refractivity (d) LogP (e) LogS and (f) LogD.

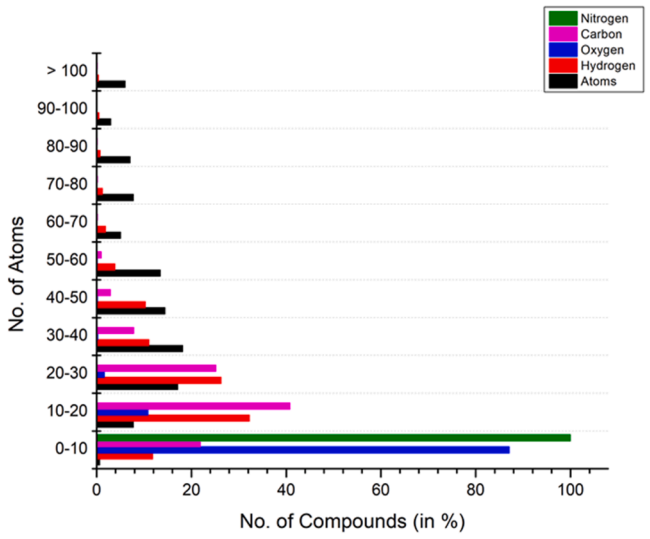


Fig. 10. Classification of phytochemical compounds with respect to the number of atoms, Nitrogen atoms, Carbon atoms, Oxygen atoms, and Hydrogen atoms where; 80–100% of compounds had between 0 and 10 atoms, respectively.

Lipinski's rule, 64 % accepted Pfizer's rule, and 52 % accepted the GSK and golden triangle rule respectively (Fig. 11).

3.1.6. ADMET

3.1.6.1. Absorption. ADMETlab 2.0 calculates seven absorption properties, out of which five parameters have been considered in this study. The probability of P-glycoprotein inhibitor and substrate was calculated and 77 % of compounds were predicted to be inhibitors while 81 % of phytochemicals have been predicted as P-glycoprotein substrates. The calculations of Human intestinal absorption (HIA) predicted about 80 % of compounds as HIA positive which indicates the phytochemicals as high absorbance levels. The human colon adenocarcinoma cell lines (Caco-2) permeability scores were also obtained and 69 % of phytochemicals were predicted to have a proper Cao-2 permeability (Fig. 12).

3.1.6.2. Distribution. The parameters to determine the phytochemicals having proper distribution pattern was also calculated. The blood-brain penetration (BBB) score was generated and it was observed that about 45 % of phytochemicals were predicted to be BBB positive. The volume distribution score revealed that 98 % of the phytochemicals were predicted to have a proper VD value (0.04–20 L/kg). The plasma protein binding (PPB) and fraction unbound (FU) scores were also calculated (Fig. 12).

3.1.6.3. Metabolism. The metabolism properties of the phytochemicals were performed to see how fast the phytochemicals can be metabolized. The scores of the isozymes responsible for the metabolism of about 80 % of the drugs were calculated and classified as inhibitors or substrates. The inhibitors and substrates of the isozymes CYP 1A2 / 2C19 / 2C9 / 2D6 / 3A4 were calculated and it was observed that the probability of the phytochemicals being a substrate is higher than the inhibitors in most isozymes except for CYP 1A2, where the probability of the compounds being an inhibitor is higher by 4 % (Fig. 12).

3.1.6.4. Excretion. The clearance (CL) and half-life (T1/2) of the compound are important pharmacokinetic parameters to determine the effectiveness of a drug. About 66 % of the total compounds were predicted to have a high clearance value ($\geq$5). For the half-life of a drug, 41 % of compounds were observed to be T1/2 positive (T1/2 $\leq$ 0.3) (Fig. 12).

3.1.6.5. Toxicity. The toxicity descriptors of the phytochemicals were calculated and considered for prediction. About 90 % of the compounds were predicted to be hERG (human ether-a-go-go-related gene) blockers. hERG plays a vital role in regulating the cardiac membrane potential waveform. The Ames test for mutagenicity was also performed and it was observed that 80 % of the phytochemicals were predicted to be Ames positive. The scores of 70 % of the phytochemicals were between

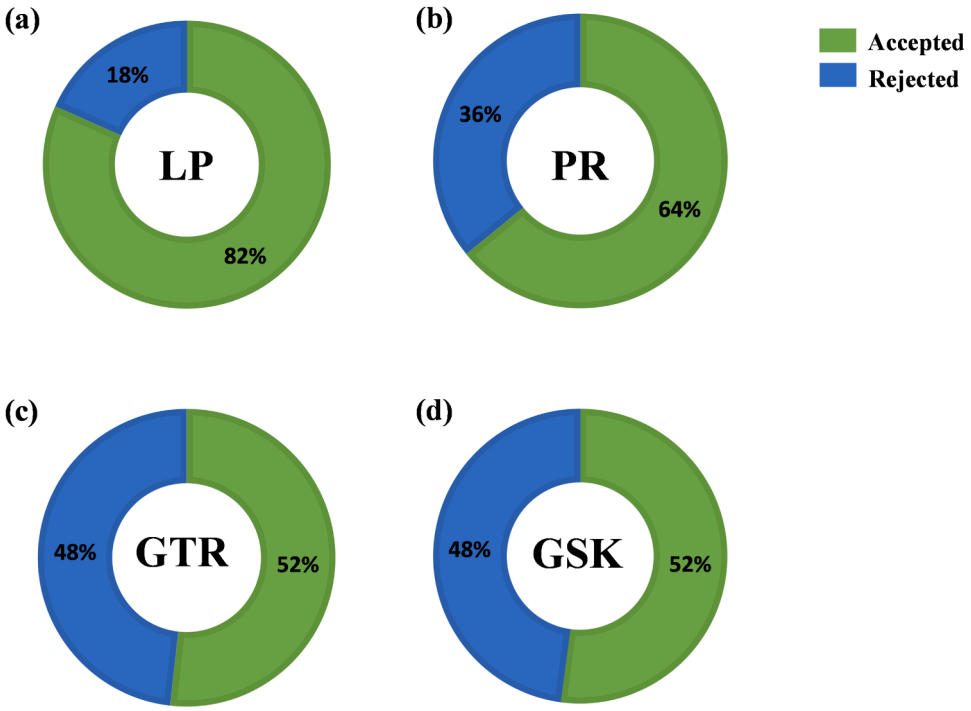


Fig. 11. Drug-likeness scores of the phytochemicals represented using four rules (a) Lipinski's Rule (b) Pfizer's Rule (c) Golden Triangle Rule, and (d) GSK Rule.

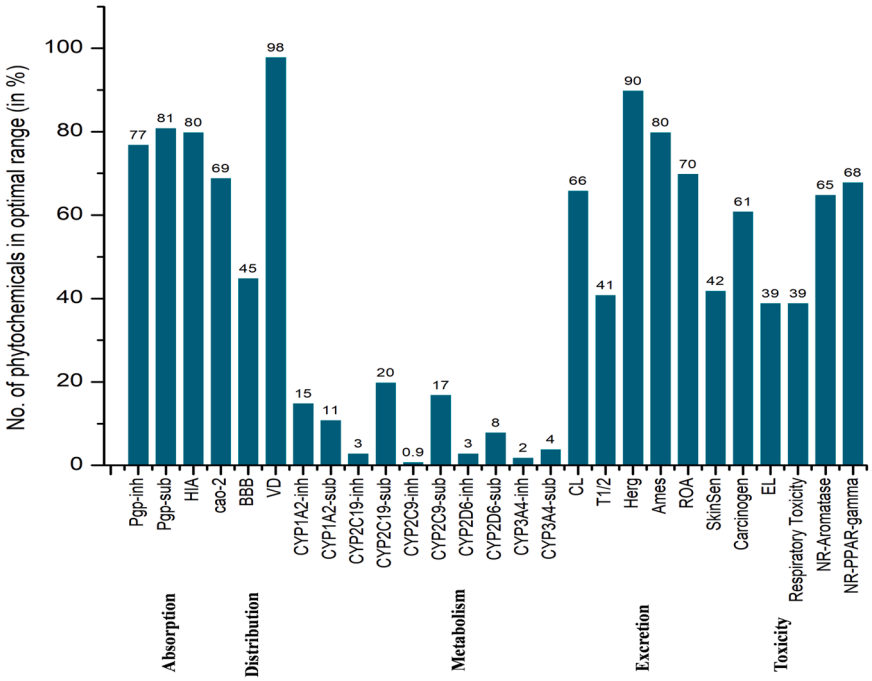


Fig. 12. Distribution of phytochemicals based on the ADMET properties.

0 and 0.3 for the rat oral acute toxicity (ROA) which indicates the probability of the compounds being toxic. The skin sensitization and carcinogenicity scores were also observed, and it was observed that 42 % and 61 % of compounds tested positive for skin sensitizer and carcinogens respectively. The other parameters also showed 40–60 % of compounds as probable toxic phytochemicals (Fig. 12).

3.1.7. Antiviral prediction

One of the unique properties of the database is the analysis of the phytochemicals to predict their antiviral potency. For this analysis, two machine learning algorithms – Random Forest and XGBoost, from (John et al., 2022) were implemented with the phytochemicals. These models were based on descriptors and were reported to have an accuracy of 88 %. The analysis showed that, almost 60 % of the compounds had the same antiviral predictions by both algorithms. For these compounds, the probability score was considered from the consensus of both models. For the remaining 40 %, the antiviral predictions were based on the highest probability scores acquired from the models. Table 3 gives the details top 20 phytochemicals with their antiviral prediction using these models sorted based on their probability scores. Fig. 13 on the other hand, shows the cumulative distributions of the phytochemicals based on their calculated probability scores for predicted to be antiviral and predicted to be not antiviral cases.

3.2. Web interface of OSADHI

OSADHI is built with a very user-friendly web interface that helps them search name of medicinal plants, state-wise availability and the associated phytochemicals. The four features of OSADHI provides easy access to traditional knowledge, geographical availability, phytochemicals and chemoinformatics reported in the database. The users can choose any of these features and find their required information. The 2D/3D structures of the phytochemicals can be downloaded into the local systems. Further, based on the classification like pathways and classes, the users can further filter their phytochemicals search. The phytochemical compounds are also linked to PubChem database for any further information. Fig. 14 shows the web interface of OSADHI with various features and search results.

4. Conclusion

OSADHI is an open-access PAN India database of medicinal plants for collating, analyzing, disseminating, integrating, and visualizing the exhaustive information of medicinal plants utilized by different ethnic communities of India. The database provides exhaustive information on the taxonomic classification, traditional knowledge and phytogeography of 6959 medicinal plants. The traditional practices documented in the database have been implemented by traditional healers for centuries and cannot be considered as scientifically valid. However, this information can be translated for the development of standardized herbal formulations, phytochemical analysis or new drug discovery after experimental validation. Extensive details of phytochemicals and their structures, physiochemical properties, ADMET, classification, and antiviral potency can be exploited as a cohesive resource by students and scientists, policymakers, and indigenous communities, as well as pharmaceutical industries in drug discovery to explore the therapeutic potential of medicinal plants against various pharmacological targets.

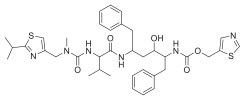
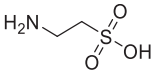
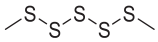
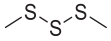
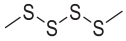
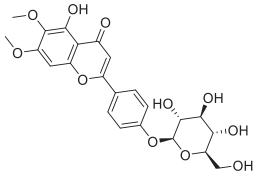
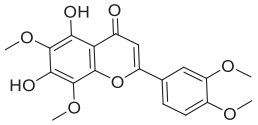
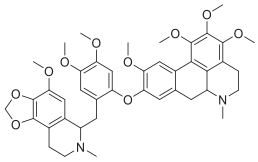
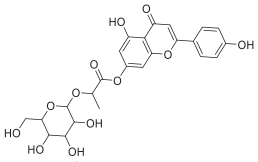
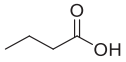
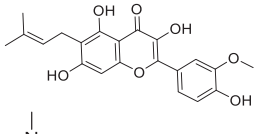
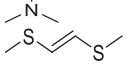
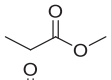
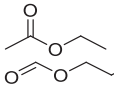
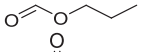
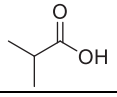
Future prospective

OSADHI aims to provide a single platform for students, researchers and other scientific communities in areas like pharmacology, drug discovery, ethno-pharmacology etc. The biological evaluation such as the in-vitro and in-vivo studies of the documented medicinal plants may be incorporated into this database. The traditional knowledge comprised in this database will also be very helpful for the scientific communities of multidisciplinary research such as herbal formulations, new pharmacophore identification, new target identification, polypharmacology, and phytopharmaceuticals.

Funding

GNS thanks DST for the financial assistance in the form of J C Bose National fellowship. KK thanks UGC NET-SRF. EJ thanks DST for INSPIRE fellowship. DBT is thanked for the financial support in the form of Centre of Excellence in Advanced Computation and Data Sciences (Ref. No: BT/PR40188/BTIS/137/27/2021).

Table 3
Top 20 phytochemicals in OSADHI with antiviral property prediction using Machine Learning models.

S. No.	Name	IUPAC Name	Predicted Antiviral	Probability scores	Structure
1.	Ritonavir Impurity O	1,3-Thiazol-5-ylmethyl N-[3-hydroxy-5-[[3-methyl-2-[[methyl-[(2-propan-2-yl-1,3-thiazol-4-yl)methyl]carbamoyl]amino]butanoyl]amino]− 1,6-diphenylhexan-2-yl]carbamate	Yes	0.992	
2.	Taurine	2-Aminoethanesulfonic acid	No	0.991	
3.	Dimethyl pentasulfide	(Methylpentasulfanyl)methane	No	0.991	
4.	Dimethyl trisulfide	(Methyltrisulfanyl)methane	No	0.990	
5.	Dimethyl thiosulfinate	Methylsulfinylsulfanylmethane	No	0.990	
6.	Dimethyl tetrasulfide	(Methyltetrasulfanyl)methane	No	0.989	
7.	Cirsimarin	5-Hydroxy-6,7-dimethoxy-2-[4-[(2 S,3 R,4 S,5 S,6 R)− 3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxyphenyl]chromen-4-one	Yes	0.988	
8.	S-Methyl methanethiosulfonate	Methylsulfonylsulfanylmethane	No	0.988	
9.	Methyl mesylate	Methyl methanesulfonate	No	0.988	
10.	Hymenoxin	2-(3,4-Dimethoxyphenyl)− 5,7-dihydroxy-6,8-dimethoxychromen-4-one	Yes	0.988	
11.	–	6-[[4,5-Dimethoxy-2-[(1,2,3,10-tetramethoxy-6-methyl-5,6,6a,7-tetrahydro-4 H-dibenzo[de,g]quinolin-9-yl)oxy]phenyl]methyl]− 4-methoxy-7-methyl-8,9-dihydro-6 H-[1,3]dioxolo[4,5-*f*]isoquinoline	Yes	0.987	
12.	–	[5-Hydroxy-2-(4-hydroxyphenyl)− 4-oxochromen-7-yl] 2-[3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxypropanoate	Yes	0.987	
13.	Butanoic acid	Butanoic acid	No	0.986	
14.	3′-O-Methylgancaonin P	3,5,7-Trihydroxy-2-(4-hydroxy-3-methoxyphenyl)− 6-(3-methylbut-2-enyl)chromen-4-one	Yes	0.986	
15.	Trimethylamine	N,N-Dimethylmethanamine	No	0.985	
16.	Bis(methylthio)ethene	1,2-Bis(methylsulfanyl)ethene	No	0.985	
17.	Methyl propanoate	Methyl propanoate	No	0.985	
18.	Ethyl acetate	Ethyl acetate	No	0.985	
19.	Propyl formate	Propyl formate	No	0.985	
20.	Isobutyric acid	2-Methylpropanoic acid	No	0.985	

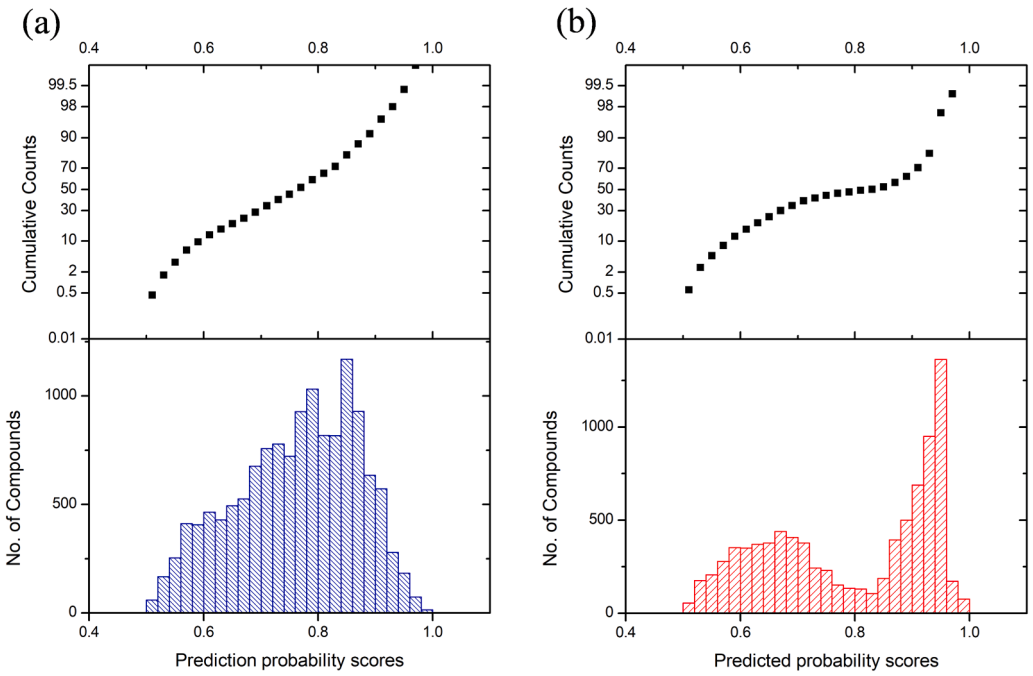


Fig. 13. Cumulative count and no. of compounds predicted of the have presence/absence of antiviral potency using machine learning models (Random Forest and XGBoost) - (a) 60 % of compounds have antiviral potency and (b) 40 % of compounds showed no antiviral potency.

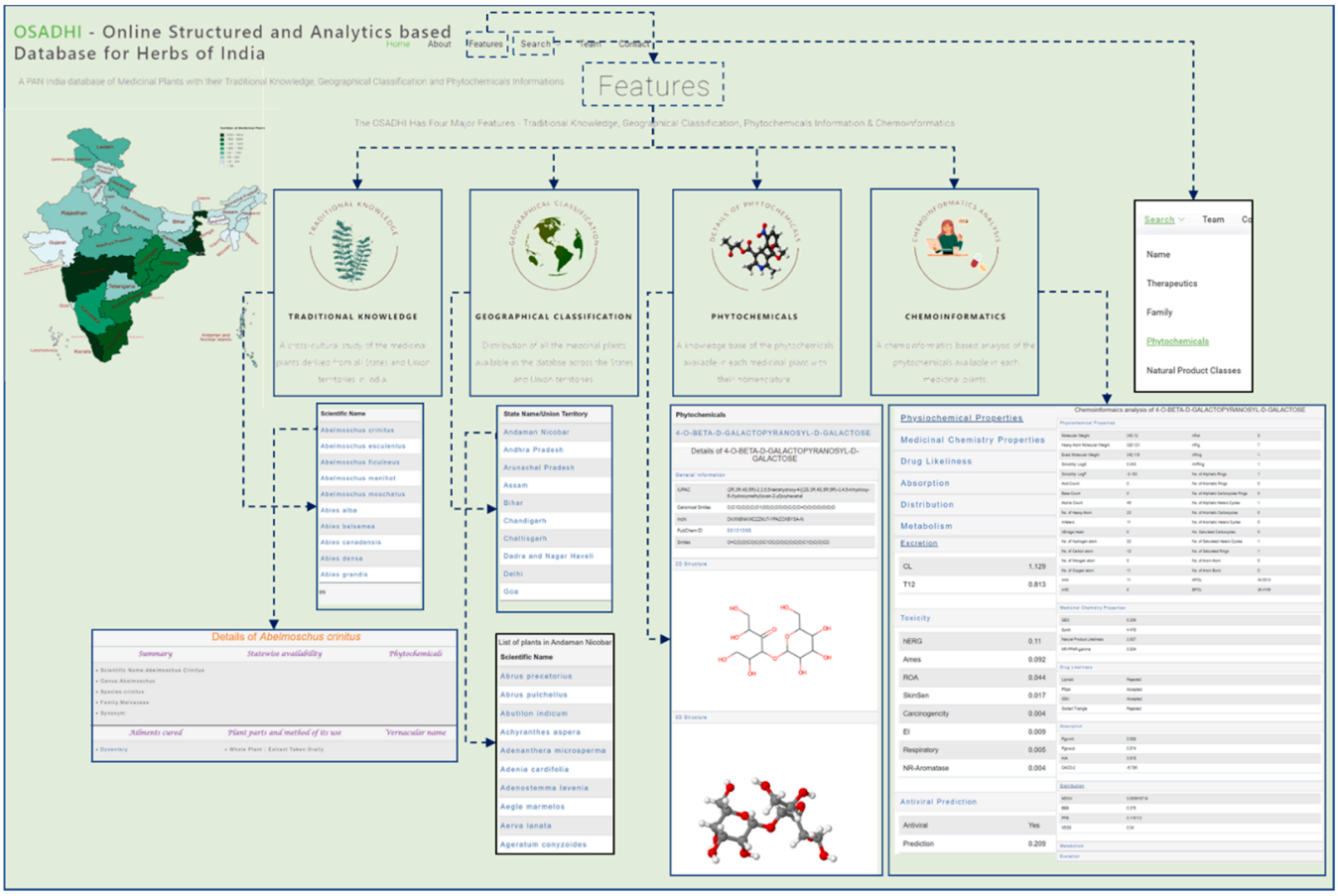


Fig. 14. Web interface of OSADHI with various features and search results (Publicly available at https://www.neist.res.in/osadhi).

CRediT authorship contribution statement

Kikrusenuo Kiewhuo: Data curation, Formal analysis, Writing – original draft. **Dipshikha Gogoi:** Data curation. **Hridoy Jyoti Mahanta:** Data curation, Formal analysis, Validation, Writing – original draft, Writing – review & editing. **Ravindra K. Rawal:** Data curation, Formal analysis, Validation. **Debabrata Das:** Data curation, Methodology. **Vaikundamani S:** Validation, Data curation. **Esther Jamir:** Data curation. **G. Narahari Sastry:** Conceptualization, Methodology, Validation, Writing – review & editing.

Conflict of interest

The authors have no conflict of interests.

References

Aslam, M.S., Ahmad, M.S., 2016. Worldwide importance of medicinal plants: current and historical perspectives. Recent Adv. Biol. Med. 2, 88–93. https://doi.org/10.18639/RABM.2016.02.338811.

Bansal, R., Jat, R.S., Kumbhani, S., Rathod, J.H., 2016. Ethnomedicinal survey of medicinal plants use from Narmada, Gujarat, India. Med. Plants Int. J. Phytomed. Relat. Ind. 8 (3), 233–237. https://doi.org/10.5958/0975-6892.2016.00021.6.

Chen, X., Li, H., Tian, L., Li, Q., Luo, J., Zhang, Y., 2020. Analysis of the physicochemical properties of acaricides based on Lipinski's rule of five. J. Comput. Syst. Biol. 27 (9), 1397–1406. https://doi.org/10.1089/cmb.2019.0323.

Dey, A., De, J.N., 2012. Ethnobotanical survey of Purulia district, West Bengal, India for medicinal plants used against gastrointestinal disorders. J. Ethnopharmacol. 143 (1), 68–80. https://doi.org/10.1016/j.jep.2012.05.064.

Fan, Jianghong, Lannoy, Ines A.M. de, 2014. Pharmacokinetics. Biochem. Pharmacol. 87 (1), 93–120. https://doi.org/10.1016/j.bcp.2013.09.007.

Fitzgerald, M., Heinrich, M., Booker, A., 2020. Medicinal plant analysis: a historical and regional discussion of emergent complex techniques. Front. Pharmacol. 10, 1480. https://doi.org/10.3389/fphar.2019.01480.

Gola, J., Obrezanova, O., Champness, E., Segall, M., 2006. ADMET property prediction: the state of the art and current challenges. QSAR Comb. Sci. 25 (12), 1172–1180. https://doi.org/10.1002/qsar.200610093.

Harvey, A.L., 2008. Natural products in drug discovery. Drug Discov. Today 13 (19–20), 894–901. https://doi.org/10.1016/j.drudis.2008.07.004.

Harvey, A.L., Edrada-Ebel, R., Quinn, R.J., 2015. The re-emergence of natural products for drug discovery in the genomics era. Nat. Rev. Drug Discov. 14 (2), 111–129. https://doi.org/10.1038/nrd4510.

Hur, J., Wild, D.J., 2008. PubChemSR: a search and retrieval tool for PubChem. Chem. Cent. J. 2 (1), 1–7. https://doi.org/10.1186/1752-153X-2-11.

Jain, A., Katewa, S.S., Galav, P.K., Sharma, P., 2005. Medicinal plant diversity of Sitamata wildlife sanctuary, Rajasthan, India. J. Ethnopharmacol. 102 (2), 143–157. https://doi.org/10.1016/j.jep.2005.05.047.

John, L., Soujanya, Y., Mahanta, H.J., Narahari Sastry, G., 2022. Chemoinformatics and machine learning approaches for identifying antiviral compounds. Mol. Inform. 41 (4), 2100190 https://doi.org/10.1002/minf.202100190.

Kar, S., Leszczynski, J., 2020. Open access in silico tools to predict the ADMET profiling of drug candidates. Expert Opin. Drug Deliv. 15 (12), 1473–1487. https://doi.org/10.1080/17460441.2020.1798926.

Kiewhuo, K., Gogoi, D., Mahanta, H.J., Rawal, R.K., Das, D., Sastry, G.N., 2022. North East India medicinal plants database (NEI-MPDB). Comput. Biol. Chem. 100, 107728 https://doi.org/10.1016/j.compbiolchem.2022.107728.

Kim, H.W., Wang, M., Leber, C.A., Nothias, L.F., Reher, R., Kang, K.B., Cottrell, G.W., 2021. NPClassifier: a deep neural network-based structural classification tool for natural products. J. Nat. Prod. 84 (11), 2795–2807. https://doi.org/10.1021/acs.jnatprod.1c00399.

Krishna, G.L., 2022. Ayurveda awaits a new dawn. Indian J. Med. Ethics 2, 1–4. https://doi.org/10.20529/IJME.2021.093.

Kumar, A., 2020. Phytochemistry, pharmacological activities and uses of traditional medicinal plant Kaempferia galanga L.–An overview. J. Ethnopharmacol. 253, 112667 https://doi.org/10.1016/j.jep.2020.112667.

Kumar, A., Kumar, R., Sharma, M., Kumar, U., Gajula, M.P., Singh, K.P., 2018. Uttarakhand medicinal plants database (UMPDB): a platform for exploring genomic, chemical, and traditional knowledge. Data 3 (1), 7. https://doi.org/10.3390/data3010007.

Kumar, Y., Prakash, O., Tripathi, H., Tandon, S., Gupta, M.M., Rahman, L.U., Khan, F., 2018. AromaDb: a database of medicinal and aromatic plant's aroma molecules with phytochemistry and therapeutic potentials. Front. Plant Sci. 9, 1081. https://doi.org/10.3389/fpls.2018.01081.

Mahomoodally, M.F., Protab, K., Aumeeruddy, M.Z., 2019. Medicinal plants brought by Indian indentured immigrants: a comparative review of ethnopharmacological uses between Mauritius and India. J. Ethnopharmacol. 234, 245–289. https://doi.org/10.1016/j.jep.2019.01.012.

Manhã, E.M., Silva, M.C., Alves, M.G., Almeida, M.B., Brandão, M.G., 2008. PLANT: a bibliographic database about medicinal plants. Rev. Bras. Farmacogn. 18, 614–617. https://doi.org/10.1590/S0102-695×2008000400020.

Mary, J.A., Priyadharshini, K.C., Amal, G.P.R., Ramya, G., Nithya, R., Ambika, M.B., Shenbagarathai, R., 2012. MEDDB: a medicinal plant database developed with the information gathered from tribal people in and around Madurai, Tamil Nadu. Bioinformation 8 (8), 391–393. https://doi.org/10.6026/97320630008391.

Meetei, P.A., Singh, P., Nongdam, P., Prabhu, N.P., Rathore, R.S., Vindal, V., 2012. NeMedPlant: a database of therapeutic applications and chemical constituents of medicinal plants from north-east region of India. Bioinformation 8 (4), 209–211. https://doi.org/10.6026/97320630008209.

Moriwaki, H., Tian, Y.S., Kawashita, N., Takagi, T., 2018. Mordred: a molecular descriptor calculator. J. Cheminform. 10 (1), 1–14. https://doi.org/10.1186/s13321-018-0258-y.

Mukherjee, P.K., Harwansh, R.K., Bahadur, S., Banerjee, S., Kar, A., Chanda, J., Katiyar, C.K., 2017. Development of Ayurveda–tradition to trend. J. Ethnopharmacol. 197, 10–24. https://doi.org/10.1016/j.jep.2016.09.024.

Ningthoujam, S.S., Talukdar, A.D., Potsangbam, K.S., Choudhury, M.D., 2012. Challenges in developing medicinal plant databases for sharing ethnopharmacological knowledge. J. Ethnopharmacol. 141 (1), 9–32. https://doi.org/10.1016/j.jep.2012.02.042.

Parthiban, R., Vijayakumar, S., Prabhu, S., Yabesh, J.G.E.M., 2016. Quantitative traditional knowledge of medicinal plants used to treat livestock diseases from Kudavasal taluk of Thiruvarur district, Tamil Nadu, India. Rev. Bras. Farmacogn. 26 (1), 109–121. https://doi.org/10.1016/j.bjp.2015.07.016.

Pathania, S., Ramakrishnan, S.M., Bagler, G., 2015. Phytochemica: a platform to explore phytochemicals of medicinal plants. Database 2015. https://doi.org/10.1093/database/bav075.

Pešić, M., Stanković, S., 2015. Development of natural product drugs in a sustainable manner. Brief for United Nations Global Sustainable Development Report.

Pouliot, M., 2011. Relying on nature's pharmacy in rural Burkina Faso: empirical evidence of the determinants of traditional medicine consumption. Soc. Sci. Med. 73 (10), 1498–1507. https://doi.org/10.1016/j.socscimed.2011.08.037.

Prabhu, S., Vijayakumar, S., Yabesh, J.M., Ravichandran, K., Sakthivel, B., 2014. Documentation and quantitative analysis of the local knowledge on medicinal plants in Kalrayan hills of Villupuram district, Tamil Nadu, India. J. Ethnopharmacol. 157, 7–20. https://doi.org/10.1016/j.jep.2014.09.014.

Ram, T.S., Munikumar, M., Raju, V.N., Devaraj, P., Boiroju, N.K., Hemalatha, R., Dhiman, K.S., 2022. In silico evaluation of the compounds of the ayurvedic drug, AYUSH-64, for the action against the SARS-CoV-2 main protease. J. Ayurveda Integr. Med. 13 (1), 100413 https://doi.org/10.1016/j.jaim.2021.02.004.

Reddy, K.N., Trimurthulu, G., Reddy, C.S., 2010. Medicinal plants used by ethnic people of Medak district, Andhra Pradesh. Ind. J. Tradit. Knowl. 9 (1), 184–190.

Saldívar-González, F.I., Lenci, E., Trabocchi, A., Medina-Franco, J.L., 2019. Exploring the chemical space and the bioactivity profile of lactams: a chemoinformatic study. RSC Adv. 9 (46), 27105–27116. https://doi.org/10.1039/C9RA04841C.

Sargia, B., Singh, B., Gupta, N., Gahlot, L.K., Gulati, T., Hasija, Y., 2018. MEDPDB: an online database of medicinal plants. J. Adv. Pharm. Educ. Res. 7 (4), 204–207.

Savithramma, N., Yugandhar, P., Suhrulatha, D., 2015. Traditional medicinal plants used by local people of Kailasakona-A sacred grove of Chittoor District, Andhra Pradesh, India. Defenders 16, 18.

Scotti, L., Scotti, M.T., 2019. Studies of ADMET properties in medicinal chemistry–part-I. Curr. Top. Med. Chem. 19 (29), 2641–2642. https://doi.org/10.2174/1568026619291911211093403.

Scotti, L., Scotti, M.T., Muratov, E., 2021. Natural products databases for drug design. Mini Rev. Med. Chem. 21 (18), 2628–2629. https://doi.org/10.2174/1389557521182102409523 6.

Sharma, A.K., Bajpai, S., Shrivastava, S., Kanungo, V.K., 2014. Inventorying medicinal plants in urban homegardens of Raipur, Chhattisgarh. Int. J. Herb. Med. 2 (1), 43–50.

Shukla, G., Chakravarty, S., 2012. Ethnobotanical plant use of Chilapatta reserved forest in West Bengal. Indian 138 (12), 1116.

Sinha, S., Hazarika, A., Johari, S., Neog, B., Rajkhowa, S., Biswas, A., 2022. IMPDB: Indian medicinal phytochemical database curated for drug designing. J. Comput. Biophys. Chem. 21 (6), 709–728. https://doi.org/10.1142/S2737416522500302.

Singh, S., Kumar, V., Gill, J.P.K., Datta, S., Singh, S., Dhaka, V., Singh, J., 2020. Herbicide glyphosate: toxicity and microbial degradation. Int. J. Environ. Res. Public Health 17 (20), 7519. https://doi.org/10.3390/ijerph17207519.

Singh, S., Singh, D.B., Singh, S., Shukla, R., Ramteke, P.W., Misra, K., 2019. Exploring medicinal plant legacy for drug discovery in post-genomic era. Proc. Natl. Acad. Sci., India Sect. B Biol. Sci. 89 (4), 1141–1151. https://doi.org/10.1007/s40011-018-1013-x.

Skoczen, S., Bussmann, R.W., 2006. ebDB-Filling the gap for an International Ethnobotany Database. Lyonia 11 (2), 71–81.

Srivastava, T.N., Rajasekharan, S., Badola, D.P., Shah, D.C., 1986. An index of the available medicinal plants used in Indian system of medicine from Jammu and Kashmir state. Anc. Sci. Life 6 (1), 49.

Thomford, N.E., Senthebane, D.A., Rowe, A., Munro, D., Seele, P., Maroyi, A., Dzobo, K., 2018. Natural products for drug discovery in the 21st century: innovations for novel drug discovery. Int. J. Mol. Sci. 19 (6), 1578. https://doi.org/10.3390/ijms19061578.

Ugboko, H.U., Nwinyi, O.C., Oranusi, S.U., Fatoki, T.H., Omonhinmin, C.A., 2020. Antimicrobial importance of medicinal plants in Nigeria. Sci. World J. 2020. https://doi.org/10.1155/2020/7059323.

Vinayak, U., Hegde, H.V., Divakar, M., Shankar, B., Veeresh, H., Kholkute, S.D., 2010. Digital database on ethno-medicinal plants of Western Ghats. Curr. Sci. 99 (12), 1658–1659.

Vivek-Ananth, R.P., Mohanraj, K., Sahoo, A.K., Samal, A., 2022. IMPPAT 2.0: an enhanced and expanded phytochemical atlas of Indian medicinal plants. bioRxiv. https://doi.org/10.1101/2022.06.17.496609.

Xiong, G., Wu, Z., Yi, J., Fu, L., Yang, Z., Hsieh, C., Cao, D., 2021. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res. 49 (W1), W5–W14. https://doi.org/10.1093/nar/gkab255.

Xu, T., Chen, W., Zhou, J., Dai, J., Li, Y., Zhao, Y., 2020. NPBS database: a chemical data resource with relational data between natural products and biological sources. Database 2020. https://doi.org/10.1093/database/baaa102.