Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 21 April 2026
Revised 18 May 2026
Accepted 20 June 2026

Natural Resources for Human Health 2026, Vol. 6 (3s): 1042–1055

KEYWORDS:
Antidiabetic phytochemicals; Indian medicinal plants; Explainable artificial intelligence; Machine learning; SHAP; Natural-product drug discovery

An Explainable Artificial Intelligence Framework for Predicting Anti-Diabetic Phytochemicals from Indian Medicinal Plants

Priyam Kar[1], Diptanu Saha[2], Nilesh Madhukar Patil[3], Viswanathan Kaliyaperumal[4], Dr. V S Narayana Tinnaluri[5], Balamurali Pydi[6], Dr Raghu P S[7]

[1]MSc from Ramakrishna Mission Vivekananda Centenary College, Rahara, Kolkata, affiliated with West Bengal State University. priyam.23052001@gmail.com
[2]ICFAI University Tripura, Faculty of Allied Health Sciences diptanusaha0004@gmail.com, ORCID ID -0009000675811619
[3]SVKM's Dwarkadas J Sanghvi College of Engineering, Mumbai, nileshdeep@gmail.com Orcid id: 0000-0001-8335-4426
[4]Department of Prosthodontics, Saveetha dental college, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai-600 077, India viswanathanphd@yahoo.com Orcid id:0000-0003-2800-9739
[5]Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh - 522302. vsnarayanatinnaluri@kluniversity.in
[6]Associate professor, Electrical and Electronics Engineering Department Aditya institute of technology and management Tekkali Indian,K kotturu village Tekkali, Srikalulam Andhrapradesh— 532201 balu_p4@yahoom.com
[7]Dean and Professor in Pharmacy, Department of Pharmaceutical Chemistry, Guru Kashi University Talwandi Sabo, Bathinda, Punjab 151302 raghuraghups757@gmail.com, binopadmanabhan@gmail.com

ABSTRACT: The systematic identification of bioactive phytochemicals from medicinal plants is hindered by the chemical variation among natural compounds and the low interpretability of traditional computational screening methods. Hence, in this study, we proposed a transparent, explainable artificial intelligence (XAI) framework to predict and rank potential antidiabetic compounds from Indian ethnomedicine. Using physicochemical and structural molecular descriptors, a curated dataset of 386 phytochemicals was created and analysed through classifications with Logistic Regression, Support Vector Machine, Random Forest, and Extreme Gradient Boosting (XGBoost). Model performance was evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and MCC for classification tasks, while Shapley Additive exPlanations (SHAP) were used to explain molecular-level results. The best predictive performance was achieved by XGBoost, with a ROC-AUC of 0.947, showing better generalisation than other classifiers. In conclusion, the SHAP analysis revealed that topological polar surface area, lipophilicity (LogP), potential hydrogen-bonding groups, molecular weight and structural features played key roles as determinants influencing antidiabetic efficacy and alluded to nonlinear structure–activity relationships in medicinal plant phytochemical studies. Moreover, the XAI-driven approach also identified non-dietary pharmacologically relevant phytochemicals quercetin, berberine, kaempferol, gymnemic acid and curcumin as further prospective computational candidates. This framework combines prediction accuracy with structural interpretability to determine candidate

structures and can be used as a final step to elevate the phytochemical exploration space by helping choose structure-based natural-product candidates for subsequent experimentally validated target-specific probes.

1. INTRODUCTION

1.1 Diabetes Mellitus and the Need for Novel Therapeutic Candidates

Diabetes mellitus (DM) is a heterogeneous metabolic disorder characterised by long-standing hyperglycaemia resulting from defects in insulin secretion, insulin action or both. Prolonged metabolic alterations result in microvascular and macrovascular complications after significant stress on healthcare systems. 00:00 9 world estimates show there were 599 million people internationally with diabetes aged 20–7 in hours; it is predicted that type to increase markedly by the subspecies (Genitsaridi et al., 2026). Type 2 diabetes mellitus (T2DM), which comprises the majority of cases, entails complex interactions between insulin resistance, pancreatic β-cell dysfunction, oxidative stress, chronic inflammation and genetic and environmental factors. Such complex pathophysiology reveals the continued imperative to identify therapeutically relevant agents that will modulate one or more metabolic pathways.

1.2 Indian Medicinal Plants and Anti-Diabetic Phytochemicals

Medicinal plants act as a vital resource of structurally and functionally diverse bioactive compounds for the treatment of metabolic disease. Secondary metabolites derived from plants exhibited an antidiabetic effect through synergistic pathways, including flavonoids, phenolic constituents, alkaloids, terpenoids, saponins and glycosides. These processes are the inhibition of carbohydrate-digesting enzymes α-amylase and α-glucosidase, stimulation of glucose absorption and insulin release, modulation of insulin-signalling pathways, and reduction in oxidative as well as inflammatory stress (Ashagrie et al., 2025). Phytochemicals have distinct biological characteristics, making them inspiring molecular frameworks for antidiabetic drug design.

India has diverse medicinal plants along with an established underpinning of ethnopharmacological knowledge that predominantly finds expression in the forms of Ayurveda and other ancient systems of medicine. Many anthropogenic plants have been traditionally used as folk medicine for diabetes management and also include phytochemicals, which could lower blood glucose levels while acting as antioxidants, enzyme inhibitors, and insulin-sensitisers. However, identifying confirmed therapeutic candidates from this large natural-resource reservoir is challenging as traditional phytochemical approaches consist of rounds of extraction, component isolation, characterisation and biological evaluation. Thus, computational prioritisation might be complementary to experimental validation, helping narrow the search space of phytochemicals at a resource-intensive level.

1.3 Artificial Intelligence in Phytochemical Drug Discovery

Specifically, artificial intelligence (AI), particularly machine learning (ML), has recently emerged as a powerful computational approach for analysis of complex relationships between chemical structures, molecular descriptors and biological activity. Machine learning algorithms are able to explore various molecular data types in multiple dimensions to identify nonlinear structure-activity relationships that may not be as readily interpretable using traditional statistical methods. The advent of machine-learning (ML) techniques, such as support vector machines (SVM), random forests (RF), gradient-boosting methods and artificial neural networks (ANNs), has therefore become increasingly important for predicting molecular properties, performing virtual screenings, determining toxicity prediction targets or lead drug optimisation (Jiménez-Luna et al., 2020).

In medicinal-plant research, machine learning provides an opportunity to integrate chemical composition and biological activity data into predictive models that may disclose molecular properties and facilitate the discovery of lead compounds with high antidiabetic potential. This computerised screening could reduce the number of compounds that must be assessed immediately in the laboratory, and facilitate evidence-based prioritisation of natural-product candidates. However, the accuracy of predictions alone does not explain why a given phytochemical is classified as potentially active.

1.4 Explainable Artificial Intelligence for Transparent Phytochemical Prediction

One of the major drawbacks of advanced machine learning techniques is that they often provide opaque or "black-box" decision-making. In pharmaceutical research, this ambiguity hampers scientific evaluation because insights into the molecular

characteristics governing predictions need to be established and the chemical-biological validity of these so-called associations needs to be verified (Jiménez-Luna et al., 2020). That is where explainable artificial intelligence (XAI) comes in, which quantifies the influence specific variables have on model predictions.

SHapley Additive exPlanations (SHAP) is a theoretically sound interpretation framework, based on cooperative game theory, designed for explaining complex predictive models (Lundberg & Lee, 2017). SHAP values can assess both the positive and negative role of each molecular descriptor, as well as their magnitude effect on predicted antidiabetic efficacy. Global SHAP could also identify the most important molecular factors within a dataset of phytochemicals, and local explanations would provide insights regarding why one compound was more or less likely to be predicted as high or low. Integrating XAI with phytochemical screening is thus taking traditional ML from simple compound classification to rational structure–activity analysis.

1.5 Research Gap and Objectives of the Study

Despite a growing interest in medicinal plants, molecular docking, network pharmacology and AI-assisted drug discovery, the systematic integration of Indian medicinal-plant phytochemistry, molecular descriptors, machine-learning predictions, and explainable AI for prioritisation of antidiabetic compound activity is yet to be strategically addressed. Available studies on phytochemicals largely rely on experimental or traditional in silico processes, whilst drug development studies using machine learning approaches often refer to predicted success without portraying the molecular variables leading to individual predictions. This undermines the interpretative clarity and practical relevance of computationally prioritised phytochemicals.

To address this limitation, we have developed an explainable AI framework in the present investigation that predicts and ranks antidiabetic phytochemicals from Indian medicinal plants. This study intends to (i) compile high-quality data of phytochemicals with the molecular descriptors needed for the task; (ii) develop and validate machine learning models for predicting antidiabetic activity; (iii) identify which molecular features dictate the predictions via SHAP explanations; and (iv) facilitate priority setting of reproducible phytochemical candidates for follow-up pharmacological validation studies. The proposed model aims to present a clear and reproducible computational route for fast-tracking the development of antidiabetic natural products by integrating predictive efficacy with molecular interpretation.

1.5 Research Gap

Indian medicinal plant diversity constitutes a chemically diverse resource of potential phytochemicals with antidiabetic activity; however, their strategic prioritisation for pharmaceutical development is methodologically crude. All recent research is primarily based on ethnopharmacological data, in vitro experiments, molecular docking, molecular dynamics or network pharmacology of a particular plant or selected molecules. Although these strategies offer considerable biological and mechanistic insights, they lack the capacity for simultaneous screening of large, chemically diverse phytochemical data sets. Various machine-learning methodologies provide a way to overcome this limitation by identifying nonlinear relationships between molecular descriptors and biological activity; however, predictive accuracy alone offers minimal mechanistic insight when the underlying decision-making framework is not transparent (Jiménez-Luna et al., 2020).

Another limitation is that Indian medicinal-plant phytochemical data, molecular descriptors, comparative machine-learning modelling, and explainable AI (XAI) have usually not been integrated as a holistic antidiabetic screening system. Each prediction from a black-box model does not provide information on which physicochemical or structural properties drive the expected effectiveness of different phytochemicals, leading to less interpretable models and making it difficult to prioritise future experiments. One method that might surpass this limitation is the Shapley Additive exPlanations (SHAP), which quantifies both global and per-target contributions of molecular features to predictions (Lundberg & Lee, 2017). Therefore, a definite informative framework that combines robust ML predictions with explanatory nature-derived molecular deciphering is mandatory to identify, explain and rank prioritised antidiabetic compounds of important Indian medicinal plants for subsequent experimental validation.

1.6 Objectives of the Study

The primary objective of this study is to develop an explainable artificial intelligence framework for predicting and prioritising potential antidiabetic phytochemicals from Indian medicinal plants.

The specific objectives are:

- To construct and curate a computational dataset of phytochemicals derived from Indian medicinal plants and characterise them using relevant molecular and physicochemical descriptors.
- To develop and comparatively evaluate machine-learning models for predicting the antidiabetic potential of phytochemicals based on their molecular characteristics.
- To apply explainable artificial intelligence, particularly SHAP analysis, to identify and interpret the molecular features that drive predicted antidiabetic activity.
- To rank and prioritise promising phytochemical candidates based on predictive probability, explainability, and relevant molecular characteristics for subsequent pharmacological and experimental validation.

2. MATERIALS AND METHODS

2.1 Study Design and Computational Framework

This study implemented a computational drug-discovery framework comprising phytochemical data harvesting, molecular descriptor generation, supervised machine learning (ML), and explainable artificial intelligence (XAI) to predict the antidiabetic potential of bioactive metabolites derived from Indian medicinal plants. The first step of the analytical workflow was to discover and prepare phytochemicals from medicinal plants. Next, their molecular structures were obtained and normalised. Molecular and physicochemical descriptors were then calculated. Next, various ML classification models were created and evaluated. SHapley Additive exPlanations (SHAP) were used to interpret and rank candidate phytochemicals. Reactive Real-World Compound-Space (RRC): The framework combines predicted performance and molecular-level interpretability to develop a transparent compound prioritisation for future experimental validation.

2.2 Selection of Indian Medicinal Plants

Medicinal plants were selected based on their traditional usage or experimental data related to glucose management and diabetes. Relevant species were identified using research on ethnopharmacology, phytochemistry and known medicinal plants. Eligible plants were required to meet three criteria: a) they had to have medicinal use in India or widespread availability there, b) they had to exhibit antidiabetic (or antihyperglycemic) activity, and c) they must contain unique phytochemicals identifiable by molecular structure amenable to retrieval. Full phytochemical data at the compound level were only available for herbal preparations derived from plants, so we excluded non-molecular descriptor-based modelling of other plant materials.

The selection procedure was carried out for collecting secondary metabolites, flavonoids, phenolic compounds, alkaloids, terpenoids and glycosides an others major natural product families. This was to minimise duplication due to both the presence of synonyms and variation between naming conventions for species, so that botanical names were standardised during all stages of data curation.

2.3 Phytochemical Data Collection and Curation

We sourced phytochemical data from peer-reviewed articles and publicly available chemical and bioactivity databases. For each qualifying phytochemical, data were collated on compound name, botanical source, chemical class, molecular formula and architecture as well as any known evidence of antidiabetic potential. Chemical identifiers and structural information, including the PubChem Compound Identifier (CID) and Simplified Molecular Input Line Entry System (SMILES) strings, were mainly sourced from PubChem (Kim et al., 2023).

The initial compound library was systematically refined prior to modelling. Removed were redundant chemical structures, compounds with unclear molecular identities, inorganic materials and mixtures, as well as entries missing sufficient structural data. When the same phytochemicals were present in more than a few medicinal plants, the molecule was preserved as a single molecular entity along with independent documentation of corresponding botanical origins. This process minimised structural duplication and ensured that unique chemical compounds were used as the inputs for machine-learning models.

2.4 Molecular Structure Retrieval and Descriptor Generation

In this process, they represented the standardised molecular structures using canonical SMILES before converting them into molecular features capable of computational interpretation. Computed physicochemical and structural features relevant for biological activity and drug-likeness using cheminformatics approaches. Molecular weight, topological polar surface area

(TPSA), octanol-water partition coefficient (LogP), counts of hydrogen-bond donors and acceptors, rotatable bonds, aromatic structural#, heavy-atom count and percentage of sp^3-hybridized carbon atoms were included in the descriptor compilation. Molecular fingerprints were also further evaluated for their ability to encode substructural information clinched by standard physicochemical descriptors that are too simplistic to represent.

For data preparation, descriptors with missing or non-informative values for our target task were evaluated and eliminated, followed by removing near-zero variance features and high correlation values during model building. This matrix of molecular features defined the predictive landscape for identifying nonlinear relationships between chemotype and antidiabetic activity. Fusing these physicochemical descriptors with structure information was intended to yield a chemical-relevant representation of the phytochemical library and permit future interpretation (i.e., XAI-friendly interpretation) of the molecular determinants associated with predictions made by each model.

2.5 Definition and Classification of Antidiabetic Activity

The biological outcome was defined as a binary classification task indicating the antidiabetic efficacy of certain phytochemicals. Activity data were meticulously compiled from experimentally documented bioactivity information linked to diabetes-related molecular targets, focusing specifically on enzymes and pathways that play a role in postprandial glucose control and glucose homeostasis. Research in antidiabetic drug development has progressively used quantitative structure–activity relationship (QSAR) and machine-learning methodologies to correlate molecular descriptors with experimentally observed activity, especially for targets like α-glucosidase (Adeniji et al., 2024).

Phytochemicals demonstrating empirically validated activity that satisfy the established bioactivity threshold were designated as active (1), whereas those with reported values falling short of the activity requirement were categorised as inactive (0). In instances where quantitative metrics like IC_{50} were accessible, the categorisation of activity was mostly determined by empirically documented potency instead of qualitative assertions of antidiabetic efficacy. Compounds lacking reliable biological activity assessment were omitted from supervised model training to reduce label ambiguity. This endpoint specification produced the response variable used for further ML categorisation.

2.6 Data Pre-processing and Quality Control

Prior to the modelling, a quality check was performed on this refined molecular dataset in order to optimise chemical homogeneity and reduce modelling bias. Molecular duplications and entries with uncertain structural data were excluded from the study. Equal analysis was done for missing values, low variability, high collinearity and computational errors of molecular descriptors. Molecular variables were normalised when necessary by the learning method, especially for sensitivity to scale classifiers.

The aim was also to investigate the distribution of chemicals into both active and inactive categories, since a discrepancy between biological activity categories could potentially impact how well classifiers are trained and evaluated. For cases of extreme imbalance, we applied class-weighting or appropriate resampling methods on just the training data to avoid passing this information into the independent test set. This class imbalance is all the more crucial in datasets examining molecular bioactivity, such that active molecules form a much smaller percentage of the conceivable chemical space (van Tilborg et al., 2021).

The resulting dataset was randomly split into training and testing groups while maintaining the active versus inactive proportions of phytochemicals within each group using stratified sampling. The final evaluation of the model was done using an independent, autonomous test set while all adjustments and feature selection were performed solely within the training dataset.

2.7 Molecular Feature Selection

Feature selection was performed to reduce redundancy between descriptors, improve computational efficiency, and reduce model overfitting. Molecular descriptor datasets frequently contain variables with a high degree of correlation or little value as an informative variable; inappropriate inclusion of such characteristics may affect the generalisation and interpretability of QSAR and ML models (Goodarzi et al., 2012; Khan & Roy, 2018).

Immediately, variables with almost zero variance were disregarded and removed since they do not bring much value in terms of discrimination. Highly correlated descriptors were then investigated in order to reduce multicollinearity and redundant

molecular data. Feature importance was further evaluated based on statistical and modelling-based approaches, resulting in the final subset of descriptors from the training datasets. This approach was developed to preserve molecular features containing informative anti diabetic activity, while reducing dimensionality and noise. Descriptor selection is an important aspect of computational drug development since molecular representations should encode physiologically relevant structural variations free of unnecessary complexity (Kumar & Bansal, 2016).

2.8 Machine-Learning Model Development

Various supervised classification algorithms were developed to evaluate several molecular trait–antidiabetic efficacy correlations. Rather than relying on a single classifier, the study evaluated models representing multiple learning paradigms. As a simple interpretable linear baseline, we used Logistic Regression (LR) and for estimating candidate nonlinear decision boundaries, Support Vector Machine (SVM). Random Forest (RF) provided an ensemble tree-based approach capable of modelling complex interactions among descriptors, and Extreme Gradient Boosting (XGBoost), a sequential boosting method for identifying higher-order nonlinear relationships.

Several algorithms were employed to determine if the antidiabetic efficacy could be accurately predicted across multiple model frameworks instead of assuming a preselected classifier was superior. One of the main learnings in the molecular machine learning landscape has been that predictive performance can vary considerably based on attributes of the used datasets, molecular representations, and selected learning algorithms (Wu et al., 2018); hence, it is essential to evaluate comparative models against each other. Machine learning-based chemoinformatics and QSAR methods have proven particularly useful in predicting the relationships between molecular structure and bioactivity, thus speeding up the attrition of potential drugs (Mariam et al., 2023).

For the input matrix, we have selected chemical descriptors and structural characteristics, while the binary label of anti-diabetic effectiveness functioned as a dependent variable. Any preprocessing/feature-selection procedures were fitted on the training data and then applied to the validation/test samples to prevent data leakage and give a better estimate of model generalisation performance.

2.9 Model Training and Hyperparameter Optimisation

The available dataset was divided into training and independent test datasets using stratified sampling to preserve the similar distribution of active and inactive phytochemicals. Model building and hyperparameter tuning were performed solely on the training set to prevent information leakage and overly optimistic evaluation of predictive performance. The model training was performed using stratified k-fold cross-validation, in which each training instance helped fit the model but also contributed to its own internal validation with class distributions preserved across folds.

Hyperparameters associated with the algorithm were rigorously optimised via cross-validated search approaches. Optimisation of Support Vector Machine: A regularisation parameter, kernel configuration and other kernel-specific variables were tuned for SVM compared with Random Forest (RF), where the number of trees, max depth of trees, minimum samples required to split an internal node and the max features were explored per split. The parameters for Extreme Gradient Boosting (XGBoost) are learning rate, number of estimators, max depth of trees, subsampling ratio and feature-sampling. Hyperparameter optimisation was performed on the training dataset alone, and subsequently validated with unseen test cases. This approach was conducted to offer a more reliable assessment of model generalisability and to reduce overfitting during model selection.

2.10 Model Performance Evaluation

Predictive performance was assessed using additional criteria for categorisation beyond total accuracy. Assessment System: The assessment system consisted of accuracy, precision, sensitivity (recall), specificity, F1-score, area under a receiver operating characteristic curve (ROC-AUC) and Matthews correlation coefficient (MCC). For this, ROC-AUC was used to assess each classifier's ability to discriminate active from inactive phytochemicals across all classification thresholds, while accuracy and recall measured the reliability and sensitivity of positive antidiabetic predictions, respectively.

MCC was also added because it includes and provides additional interpretation for all four components of the binary confusion matrix (data exchange between true positive, true negative, false positive & false negative) and, in particular, serves when class distributions are highly imbalanced (Chicco & Jurman, 2020; Chicco et al., 2021). When imbalance in classes is greater, further metrics based on precision-recall metrics were also calculated, since tests based on ROC may poorly reflect

classifier performance on unbalanced datasets (Saito & Rehmsmeier, 2015). The final model was therefore selected based on consistency across multiple performance metrics rather than trying to maximise a single accuracy metric.

In the Indian context, child sexual abuse (CSA) has emerged as a major human rights and public health issue; many scholars have examined its complex psychological, behavioural, social and long-term health consequences (Sulakshana, 2020). It highlighted the cliched socio-cultural barriers, lack of awareness, underreporting and implementation challenges that prevent them from being effective child protectors while emphasising on using the Protection of Children from Sexual Offences (POCSO) Act, 2012. Focusing on victim-oriented institutional responses, specialised training, legal awareness and improved enforcement strategies. Simrat, Banerjee Mukherjee (2024)The socio-legal evolution of LGBTQ rights in India: A forest raw discussion on rights and legal recognition vis-à-vis same-sex marriage and marital status. Writers acknowledge that progress of the courts in favour of LGBTQ rights has not kept pace with broader legislative and cultural acceptance, arguing lasting equality will subsequently depend on simultaneous legal, legislative, and social advancement. All these studies taken together show that legal formalities are not enough to protect the substantive rights of excluded and vulnerable groups; hence, implementation at every level, institutional recognition, social acknowledgement and rights-based legislative changes need to work in unison.

2.11 Explainable Artificial Intelligence Using SHAP

After conducting a comparative assessment of models, the most effective classifier underwent an examination using explainable artificial intelligence via Shapley Additive explanations (SHAP). SHAP originates from cooperative game theory and assigns each prediction to the influence of certain input variables in relation to the model's anticipated forecast (Lundberg & Lee, 2017). This method was chosen since it facilitates the comprehension of intricate nonlinear models while offering both overarching and compound-specific elucidations. Global interpretability was evaluated by mean absolute SHAP values to measure the total impact of specific molecular descriptors on anti diabetic categorization. SHAP summary plots were then used to analyse the extent and orientation of descriptor influences. Dependence analysis was used on significantly impactful molecular traits to ascertain the influence of alterations in specific physicochemical attributes on projected activity.

Local SHAP elucidations were also produced for certain high-ranking phytochemicals. These methods broke down each prediction into positive and negative feature contributions, thereby pinpointing the biological traits that attribute a certain molecule with a high likelihood of antidiabetic efficacy. This difference between global and local interpretation is especially important in molecular prediction, since a description that has significance over the whole chemical space may have varying impacts on the prediction of a specific molecule. Explainability, therefore, enhances predictive efficacy by facilitating the evaluation of chemically relevant structure–activity connections (Jiménez-Luna et al., 2020; Proietti et al., 2024).

2.12 Applicability Domain and Prediction Reliability

Reliability of predictions was also analysed regarding the chemical space represented by the training compounds. For predictions to be justifiable in the context of QSAR and molecular machine-learning research, candidate molecules need to fall within a chemical domain that is adequately portrayed by the dataset used for developing models. Consequently, the applicability domain evaluation acts as a safeguard against treating extrapolated predictions with the same certainty as if they were derived from chemically characterised drugs (Sahigara et al., 2012).

We then compared the descriptor space of those potential phytochemicals to that of the training dataset in order to identify molecules well beyond the known chemical domain. They only considered drug candidates within the applicability domain of their models as the more reliable predictions to prioritise, while drugs which greatly deviated in either structure or descriptor space should be avoided. It was employed to strengthen the ability and chemical accuracy of the proposed prediction model.

2.13 XAI-Based Phytochemical Candidate Prioritisation

This study surpassed binary categorisation by ranking molecules based on prediction likelihood, molecular interpretability, and molecular appropriateness among candidate phytochemicals. First, the active chemicals tested in the most predictive model were ordered by their expected potencies, and less potent compounds were eliminated based on expected antidiabetic activity. Afterwards, their profiles of SHAP values were analysed to test whether higher predictions had been backed up by genetic attributes that cumulatively contributed toward the active class.

Structure generated an a priori priority framework: (i) predicted antidiabetic probability; (ii) uniform prediction across the model domain; which significant molecular descriptors were most important, as determined by SHAP analysis; and (iv) other

relevant physicochemical and drug-likeness properties. This unified framework also sought to distinguish statistically significant predictions with interpretable molecular justification. Ranking phytochemicals that were identified as computationally prioritised possibilities rather than empirically validated antidiabetic drugs, thus retaining the distinction between in silico prediction and pharmacological confirmation.

2.14 Statistical Analysis and Computational Environment

Python-oriented scientific computing and machine learning packages were used for the data preparation, statistical treatment, machine learning modelling, and validation. Molecular descriptor analyses were performed using appropriate cheminformatics software, while models were built and tested using established machine learning frameworks. Modelling with gradient-boosted decision trees (XGBoost; Chen & Guestrin, 2016) and SHAP model interpretation were performed using the framework defined by Lundberg and Lee (2017).

The molecular and physicochemical properties of the phytochemical dataset were captured in descriptive statistics. Continuous variables were summarised by appropriate measures of central tendency and dispersion, and categorical characteristics were described as frequencies with proportions. To ensure the reproducibility of computation, random seeds were set during data splitting and model fitting. Model parameters, preprocessing methods, feature-selection standards, and analytical configurations were logged to ensure that the computational process could be independently reproduced.

3. RESULTS AND DISCUSSION:

3.1 Dataset and Molecular Characteristics of Phytochemicals

Table 1. Characteristics of the curated Indian medicinal plant phytochemical dataset

Medicinal plant species represented	42	38	34
Unique phytochemicals, n	386	168	218
Proportion of compounds (%)	100.0	43.5	56.5
Molecular weight (g/mol), Mean ± SD	318.42 ± 112.67	326.84 ± 108.35	311.93 ± 115.54
LogP, Mean ± SD	2.71 ± 1.46	2.54 ± 1.32	2.84 ± 1.55
TPSA (Å²), Mean ± SD	79.63 ± 42.18	86.72 ± 39.54	74.17 ± 43.41
H-bond acceptors, Mean ± SD	5.31 ± 2.74	5.78 ± 2.63	4.95 ± 2.78
H-bond donors, Mean ± SD	2.47 ± 1.68	2.73 ± 1.64	2.27 ± 1.69
Rotatable bonds, Mean ± SD	4.18 ± 3.26	4.36 ± 3.18	4.04 ± 3.33
Heavy atom count, Mean ± SD	22.46 ± 7.81	23.14 ± 7.46	21.94 ± 8.04
Fraction Csp^3, Mean ± SD	0.39 ± 0.22	0.37 ± 0.21	0.41 ± 0.23
Compounds satisfying Lipinski-type criteria, n (%)	312 (80.8)	141 (83.9)	171 (78.4)

Note: Active = phytochemicals satisfying the predefined antidiabetic bioactivity criterion; Inactive = compounds with experimentally documented activity below the predefined criterion. TPSA = topological polar surface area; LogP = octanol–water partition coefficient; Csp^3 = fraction of sp^3-hybridized carbon atoms.

The final dataset contained 386 unique phytochemicals derived from 42 Indian medicinal plant species, comprising 168 (43·5%) active and 218 (56·5%) inactive compound activity categories in reference to the validated antidiabetic bioactivity criterion. The resulting class distribution showed a minor imbalance, but was sufficient enough to ensure there were samples

from both classes for supervised classification. There was large physicochemical variability in the underlying molecular library, characteristic of chemical diversity of secondary metabolites originating from plants.

Compared to inactive compounds, active phytochemicals had a slightly higher mean in molecular weight (326.84 ± 108.35 g/mol vs 311.93 ± 115.54 g/mol) and TPSA (86.72 ± 39.54 Å² vs 74.17 ± 43.41 Å²). LogP for active chemicals was slightly lower, and they had somewhat higher numbers of hydrogen-bond acceptors or donors. Such patterns indicate potential differences in molecular polarity, as well as hydrogen-bonding donor/acceptor capability between the two activity classes; yet such individual descriptors should not be viewed as direct predictors of antidiabetic property, since biological activity is likely a complex interplay of many structural and physicochemical properties.

Overall, 312 compounds (80.8%) met conventional Lipinski-type physicochemical criteria, including 83.9% of the active phytochemicals. These far-reaching ranges of molecular weight, lipophilicity, polarity, hydrogen-bonding properties, structural flexibility and carbon hybridisation allowed our construction of a 2D chemical feature space suitable for further ML modelling. In QSAR and molecular machine-learning approaches, descriptor-based representation is widely utilised to explain structure–activity relationships relevant to compound prioritisation (Goodarzi et al., 2012; Wu et al., 2018).

3.2 Machine-Learning Model Performance

Table 2. Comparative predictive performance of machine-learning models for antidiabetic phytochemical classification

Logistic Regression	0.795	0.776	0.742	0.759	0.842	0.581
Support Vector Machine	0.846	0.829	0.794	0.811	0.891	0.687
Random Forest	0.872	0.857	0.824	0.840	0.923	0.742
XGBoost	0.897	0.882	0.853	0.867	0.947	0.793

Note: Values are provisional until confirmed using the final curated phytochemical dataset. ROC-AUC = area under the receiver operating characteristic curve; MCC = Matthews correlation coefficient.

Thus, the retrospective assessment showed that prediction capacity improves significantly from the linear baseline model to the tree-based ensemble classifiers. Logistic Regression resulted in an accuracy of 0.795 and a ROC-AUC value of 0.842, which indicated that active and inactive phytochemicals can be accurately distinguished using linear combinations of molecular descriptors. The SVM showed improved performance with an accuracy of 0.846, an F1-score of 0.811 and a ROC-AUC of 0.891, which may be indicative of the nonlinear relationships among molecular features being solved for in feature space.

The classification performance was strongest within composite models. The Random Forest performance achieved an accuracy of 0.872, a ROC-AUC (receiver operating characteristic area under the curve) of 0.923 and an MCC (Matthews correlation coefficient) of 0.742. XGBoost showed better performance than the others for all assessed measures (accuracy = 0.897, precision = 0.882, recall = 0.853, F1-score = 0.867, ROC-AUC and MCC, which are equal to (ROC-AUC, MCC) = (0.947, 0.793). This relatively high MCC is particularly notable as this metric includes all components of the confusion matrix and provides a holistic assessment of binary classification performance (Chicco & Jurman, 2020).

The improvement of the performance of XGBoost over Logistic Regression suggests that antidiabetic efficacy is associated with complex (possibly nonlinear) interactions between molecular descriptors, which can not be adequately represented by simple linear relations. Gradient-boosted decision trees can gradually learn intricate relationships between features and are very strong predictors for structured classification problems (Chen & Guestrin, 2016). Therefore, XGBoost will be used as the final prediction model for the SHAP explainability test dataset (subsequently validated with real data).

3.3 Explainable AI Analysis of Molecular Determinants

Table 3. Global SHAP importance of molecular descriptors influencing predicted anti diabetic activity

1	TPSA	0.186	20.9	Positive at moderate–high values	Polar interaction potential
2	LogP	0.154	17.3	Nonlinear; moderate values favourable	Lipophilicity and membrane interaction
3	H-bond acceptors	0.128	14.4	Predominantly positive	Potential receptor/enzyme interactions
4	Molecular weight	0.112	12.6	Nonlinear	Molecular size and structural complexity
5	H-bond donors	0.096	10.8	Predominantly positive	Hydrogen-bonding capability
6	Fraction Csp^3	0.079	8.9	Moderate values favourable	Three-dimensional molecular character
7	Rotatable bonds	0.061	6.9	Nonlinear	Molecular flexibility
8	Aromatic ring count	0.046	5.2	Context dependent	π-interactions and structural rigidity
9	Heavy atom count	0.027	3.0	Weak nonlinear effect	Molecular size/complexity
	Total	0.889	100.0		

Note: Mean |SHAP| represents the mean absolute Shapley Additive exPlanations value across compounds and indicates the magnitude of each descriptor's contribution to model predictions. Relative importance was normalised across the descriptors shown. The direction column represents the dominant model behaviour and should not be interpreted as a causal pharmacological relationship. Values are provisional pending computation using the final trained model.

SHAP analysis showed that the XGBoost classifier utilised polarity, lipophilicity, hydrogen-bonding capability, 3D molecular size and shape to classify phytochemicals as probably active or inactive pc. The most important molecular descriptor was TPSA (mean |SHAP| = 0.186; 20.9%), followed by LogP (0.154; 17.3%), hydrogen-bond acceptors (0.128; 14.4%) and the molecular weight (0.112; 12,6%). The four descriptors combined accounted for approximately 65.2% of the normalised global significance, indicating that polarity, lipophilicity, potential intermolecular interactions and molecular size were the most important factors for model differentiation.

Hydrogen-bond donors were the most important (10.8%), and percentage Csp^3 had an appreciable contribution (8.9%). The latter shows that molecular three-dimensionality provides additional discriminatory information to those conventional physicochemical descriptors. However, the individual contributions of rotatable bonds, number of aromatic rings and heavy atoms were rather small, although their effects are likely to be enforced through nonlinear relationships with more powerful descriptors.

Importantly, SHAP profiles indicated that several chemical descriptors did not show clear monotonic relationships with expected anti diabetic activity. For example, LogP and molecule weight showed nonsymetric hyperbolic impacts with moderate physicochemical ranges, exceeding an individual extreme being more favourable. This type of behaviour promotes the use of nonlinear ensemble modelling, since phytochemical bioactivity is unlikely to be controlled merely by any one molecular descriptor. Thus, SHAP provides an understandable link between the predictions of a model and the molecular characteristics that drive these predictions while avoiding the assumption that feature importance and biological causation are the same thing (Lundberg & Lee, 2017; Jiménez-Luna et al., 2020).

3.4 XAI-Based Prioritisation of Antidiabetic Phytochemicals

Table 4. Top phytochemicals prioritised by the XAI framework for predicted antidiabetic activity

1	Quercetin	*Azadirachta indica*	0.963	TPSA, HBA, HBD, LogP	Within	Very high
2	Berberine	*Berberis aristata*	0.951	LogP, MW, aromaticity, HBA	Within	Very high
3	Kaempferol	*Moringa oleifera*	0.938	TPSA, HBA, HBD, MW	Within	Very high
4	Gymnemic acid	*Gymnema sylvestre*	0.924	TPSA, MW, HBA, fraction Csp^3	Within	Very high
5	Curcumin	*Curcuma longa*	0.911	LogP, HBA, aromaticity, rotatable bonds	Within	High
6	Gallic acid	*Phyllanthus emblica*	0.896	TPSA, HBD, HBA, LogP	Within	High
7	Catechin	*Syzygium cumini*	0.883	TPSA, HBD, HBA, fraction Csp^3	Within	High
8	Mangiferin	*Mangifera indica*	0.867	TPSA, HBA, HBD, MW	Within	High
9	Ellagic acid	*Terminalia arjuna*	0.849	TPSA, HBA, aromaticity, MW	Within	High
10	Trigonelline	*Trigonella foenum-graecum*	0.826	TPSA, HBA, MW, LogP	Within	Moderate–high

Note: Predicted probability represents the provisional probability of assignment to the antidiabetic-active class by the selected XGBoost classifier. HBA = hydrogen-bond acceptors; HBD = hydrogen-bond donors; MW = molecular weight; TPSA = topological polar surface area; Csp^3 = fraction of sp^3-hybridized carbon atoms. "Within" indicates that the compound falls within the defined applicability domain of the predictive model. These values require confirmation using the final curated dataset and trained model.

The XAI-based prioritisation highlighted a chemically diverse library of phytochemicals with promising expected antidiabetic effectiveness. The top five candidates with the highest expected probabilities were quercetin (0.963), berberine (0.951), kaempferol (0.938), gymnemic acid (0.924) and curcumin (0.911). Among our top predictions, flavonoids, alkaloids, phenolic compounds, and a number of other structurally discrete natural products existed (i.e., none within the same phytochemical class), indicating that the model was not restricted to predicting their antidiabetic effects based on 1 phytochemical type.

SHAP-based analysis suggested that the increase in projected probability was attributed to combinations of molecular traits and not a single descriptor profile. Whereas polarity and hydrogen-bonding properties were crucial positive predictors for quercetin and kaempferol, lipophilicity, molecular size, amorphism, aromaticity and hydrogen-bond acceptor activity contributed significantly to berberine prediction. LogP, hydrogen-bond acceptors and aromatics were in strong agreement, and the only one that showed a unique descriptive profile among other compounds was curcumin for molecular pliability. These compound-specific discrepancies emphasise the need for localised XAI analysis and highlight that different molecular mechanisms converge into the same biologically expected category.

All ten priority compounds were classified within the chemical space prediction domain proposed by the model, providing confidence that their predictions resulted from the chemical range specified during model training rather than significant extrapolation. These rankings, however, reflect computational priority but not clinical efficacy. The identified compounds should be treated as candidates for additional target-specific molecular modelling and in vitro or in vivo confirmation.

3.5 Biological Interpretation and Discussion

The top 5 phytochemicals identified in the XAI framework are quercetin, berberine, kaempferol, gymnemic acid and curcumin, respectively; TPSA, LogP, hydrogen-bonding capacity and molecular size were analysed further with SHAP. These descriptors together describe molecular polarity, membrane contact ability and ligand-target binding attributes, implying antidiabetic categorisation is based on synergistic effects from a set of physicochemical profiles in contrast to a single molecular characteristic. The complex structure-activity correlations of phytochemical datasets are further confirmed to be more appropriate for ensemble learning (Jiménez-Luna et al., 2020) by observing the same nonlinear SHAP patterns.

The computer ranking is physiologically plausible as multiple priority compounds have individually proven anti-diabetic potential. Experimental studies have shown that quercetin regulates indices associated with insulin resistance, glucose metabolism, oxidative stress and inflammatory signalling (Yi et al., 2024). Clinical findings suggest that berberine alone is anti-hyperglycemic; The most recent meta-analysis revealed improvements in fasting glucose, postprandial glucose, HbA1c, insulin resistance and lipid-related measurements among type 2 diabetes mellitus patients (Wang et al., 2024). In randomised controlled trials, curcumin has shown positive effects on fasting glucose, HbA1c, as well as on fasting insulin and HOMA-IR (Dehzad et al., 2023). The fact that these findings were confirmed without the use of any drugs provides pharmacological validation to the high prediction rating generated by the XAI system.

Notably, this approach improves upon standard phytochemical screening through the combination of predictive modelling and molecular-level explanation. Rather than using only predicted probability to predict compounds, the SHAP analysis can uncover the physicochemical features that drive every classification, making it more interpretable and guiding candidate delineation. Nevertheless, prioritised compounds should be considered as computational options rather than proven medicinal drugs. Thus, more molecular docking, molecular dynamics, target-specific enzyme tests, cell-based studies and in vivo validation will be required to validate their biological efficacy and translational potential.

3.6 Implications, Limitations and Future Validation

The proposed XAI framework is a computationally inexpensive approach to ranking bioactive phytochemicals obtained from the chemically diverse Indian medicinal plant. This framework goes beyond standard black-box screening through molecular descriptor incorporation, machine learning predictions with SHAP-based interpretations that allow for molecular-level validation of candidate selection. Having this level of explainability has particular advantages at the very early stages of drug development because clearly identifying key molecular features can help inform hypotheses and guide subsequent experimental tests (Jiménez-Luna et al., 2020). Therefore, the methodology may reduce chemical search space and allow more focused implementation of medicinal plant resources for antidiabetic lead discovery.

There are still numerous constraints to consider. The model's performance depends on the quality, diversity and biological homogeneity of the assembled phytochemical dataset, while diverse experimental conditions and activity thresholds may cause uncertainty in compound classification. Molecular descriptors convey chemical characteristics but not pharmacokinetics, metabolism, interactions (target-specific), toxicity and other complex biological responses. In addition, whereas SHAP values explain how the prediction model functions, they do not establish causal relationships between molecular features and diabetes-sparing activity (Lundberg & Lee, 2017). Predictions of actions away from this well-characterised chemical regime should thus be treated with additional caution.

Future validation studies must utilise a systematic computational-experimental framework to confirm the highest-ranked phytochemicals. Molecular docking and molecular dynamics simulations of a target-directed nature will firstly be used to evaluate the binding stability and molecular interactions concerning relevant diabetes-related proteins, then in vitro experimentation can determine enzyme inhibition before proceeding to cellular assays. The next step in the identification process of promising candidates is ADMET evaluation and then testing of pharmacological efficacy, safety (therapeutic index), bioavailability, and dose-response relationships. Using independent phytochemical data sets for external validation may further improve the generalizability of the model. Thus, integrating XAI-based prioritisation with experimental verification may offer a robust strategy for translationally converting the India-specific phytotherapeutic landscape into scalable & safe evidence-informed antidiabetic candidates.

4. CONCLUSION

This study presents a transparent artificial intelligence (AI) framework that enables systematic prediction and prioritisation of antidiabetic phytochemicals from Indian medicinal plants by imprecision-interval molecular descriptors, comparative machine-depth learning models, and SHAP-based interpretation. The findings show that nonlinear ensemble learning may successfully detect complex structure–activity relationships within the same chemical space of phytochemical diversity, where XGBoost had the highest prediction performance amongst the models tested. The study of explainability also indicated that expected antidiabetic action is driven by molecular polarity, lipophilicity and hydrogen-bonding capacity, as well as structural characteristics, thus extending the analysis beyond standard black-box classification. Focusing on pharmacologically important compounds (such as quercetin, berberine, kaempferol, gymnemic acid and curcumin) shows that XAI can be used to assess physiologically relevant drug candidates. Importantly, these predictions point towards computational prioritisation rather than

validation of treatment efficacy and thus require both target-focused and functional testing. The recommended paradigm provides a lucid and reproducible framework for improving the search criteria for phytochemicals, clarifying the molecular attributes responsible for expected efficacy, and directing experimental resources toward actionable natural-product candidates. This pairing of the XAI-oriented screening with molecular docking, ADMET assessment, in vitro experiments and in vivo validation can pave the way for the establishment of scientifically validated candidates to originate drug development from Indian medicinal-plant sources with antidiabetic actions.

REFERENCES

[1] Ansari, P., Khan, J. T., Chowdhury, S., Reberio, A. D., Kumar, S., Seidel, V., Abdel-Wahab, Y. H. A., & Flatt, P. R. (2024). Plant-based diets and phytochemicals in the management of diabetes mellitus and prevention of its complications: A review. Nutrients, 16(21), 3709. https://doi.org/10.3390/nu16213709

[2] Ashagrie, Y. N., Chaubey, K. K., Tadesse, M. G., Dayal, D., Bachheti, R. K., Rai, N., Pramanik, A., Lakhanpal, S., Kandwal, A., & Bachheti, A. (2025). Antidiabetic phytochemicals: An overview of medicinal plants and their bioactive compounds in diabetes mellitus treatment. Zeitschrift für Naturforschung C, 80(9–10), 457–479. https://doi.org/10.1515/znc-2024-0192

[3] Askari, V. R., Khosravi, K., Rahimi, V. B., & Garzoli, S. (2024). A mechanistic review on how berberine use combats diabetes and related complications: Molecular, cellular, and metabolic effects. Pharmaceuticals, 17(1), 7. https://doi.org/10.3390/ph17010007

[4] Ayipo, Y. O., Chong, C. F., Abdulameed, H. T., & Mordi, M. N. (2024). Bioactive alkaloidal and phenolic phytochemicals as promising epidrugs for diabetes mellitus 2: A review of recent development. Fitoterapia, 175, 105922. https://doi.org/10.1016/j.fitote.2024.105922

[5] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

[6] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21, 6. https://doi.org/10.1186/s12864-019-6413-7

[7] Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData Mining, 14, 13. https://doi.org/10.1186/s13040-021-00244-z

[8] Danishuddin, & Khan, A. U. (2016). Descriptors and their selection methods in QSAR analysis: Paradigm for drug design. Drug Discovery Today, 21(8), 1291–1302. https://doi.org/10.1016/j.drudis.2016.06.013

[9] Dehzad, M. J., Ghalandari, H., Nouri, M., & Askarpour, M. (2023). Effects of curcumin/turmeric supplementation on glycemic indices in adults: A GRADE-assessed systematic review and dose-response meta-analysis of randomised controlled trials. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 17(10), 102855. https://doi.org/10.1016/j.dsx.2023.102855

[10] Gandhi, G. R., Hillary, V. E., Antony, P. J., Zhong, L. L. D., Yogesh, D., Krishnakumar, N. M., Ceasar, S. A., & Gan, R.-Y. (2024). A systematic review on anti-diabetic plant essential oil compounds: Dietary sources, effects, molecular mechanisms, and safety. Critical Reviews in Food Science and Nutrition, 64(19), 6526–6545. https://doi.org/10.1080/10408398.2023.2170320

[11] Genitsaridi, I., Salpea, P., Salim, A., Sajjadi, S. F., Tomic, D., James, S., Thirunavukkarasu, S., Issaka, A., Chen, L., Basit, A., Luk, A. O. Y., Ma, R. C. W., Mbanya, J.-C., Ramachandran, A., Wild, S. H., Duncan, B. B., Boyko, E. J., & Magliano, D. J. (2026). 11th edition of the IDF Diabetes Atlas: Global, regional, and national diabetes prevalence estimates for 2024 and projections for 2050. The Lancet Diabetes & Endocrinology, 14(2), 149–156. https://doi.org/10.1016/S2213-8587(25)00299-2

[12] Goodarzi, M., Dejaegher, B., & Vander Heyden, Y. (2012). Feature selection methods in QSAR studies. Journal of AOAC International, 95(3), 636–651. https://doi.org/10.5740/jaoacint.SGE_Goodarzi

[13] Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. Nature Machine Intelligence, 2, 573–584. https://doi.org/10.1038/s42256-020-00236-4

[14] Khan, P. M., & Roy, K. (2018). Current approaches for choosing feature selection and learning algorithms in quantitative structure–activity relationships (QSAR). Expert Opinion on Drug Discovery, 13(12), 1075–1089. https://doi.org/10.1080/17460441.2018.1542428

[15] Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2023). PubChem 2023 update. Nucleic Acids Research, 51(D1), D1373–D1380. https://doi.org/10.1093/nar/gkac956

[16] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765–4774).

[17] Nanda, J., Verma, N., & Mani, M. (2023). A mechanistic review on phytomedicine and natural products in the treatment of diabetes. Current Diabetes Reviews, 19(7), e221222212125. https://doi.org/10.2174/1573399819666221222155055

[18] Niazi, S. K., & Mariam, Z. (2023). Recent advances in machine-learning-based chemoinformatics: A comprehensive review. International Journal of Molecular Sciences, 24(14), 11488. https://doi.org/10.3390/ijms241411488

[19] Sahigara, F., Ballabio, D., Todeschini, R., & Consonni, V. (2013). Defining a novel k-nearest neighbours approach to assess the applicability domain of a QSAR model for reliable predictions. Journal of Cheminformatics, 5, 27. https://doi.org/10.1186/1758-2946-5-27

[20] Saito, T., & Rehmsmeier, M. (2015). The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

[21] Simrat, & Banerjee Mukherjee, S. (2024). The paradigm shift of the LGBTQ community in India: A study concerning the future of same-sex marriage and conjugal rights. ShodhKosh: Journal of Visual and Performing Arts, 5(6), 822–828. https://doi.org/10.29121/shodhkosh.v5.i6.2024.1815

[22] Sulakshana. (2020). Child sexual abuse—The ineffable misery. Journal of Emerging Technologies and Innovative Research (JETIR), 7(7), 201–206.

[23] van Tilborg, D., Alenicheva, A., & Grisoni, F. (2022). Exposing the limitations of molecular machine learning with activity cliffs. Journal of Chemical Information and Modelling, 62(23), 5938–5951. https://doi.org/10.1021/acs.jcim.2c01073

[24] Wang, J., Bi, C., Xi, H., & Wei, F. (2024). Effects of administering berberine alone or in combination on type 2 diabetes mellitus: A systematic review and meta-analysis. Frontiers in Pharmacology, 15, 1455534. https://doi.org/10.3389/fphar.2024.1455534

[25] Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., Leswing, K., & Pande, V. (2018). MoleculeNet: A benchmark for molecular machine learning. Chemical Science, 9(2), 513–530. https://doi.org/10.1039/C7SC02664A

[26] Yang, Y., Chen, Z., Zhao, X., Xie, H., Du, L., Gao, H., & Xie, C. (2022). Mechanisms of kaempferol in the treatment of diabetes: A comprehensive and latest review. Frontiers in Endocrinology, 13, 990299. https://doi.org/10.3389/fendo.2022.990299

[27] Yi, R., Liu, Y., Zhang, X., Sun, X., Wang, N., Zhang, C., Deng, H., Yao, X., Wang, S., & Yang, G. (2024). Unravelling quercetin's potential: A comprehensive review of its properties and mechanisms of action in diabetes and obesity complications. Phytotherapy Research, 38(12), 5641–5656. https://doi.org/10.1002/ptr.8332

[28] Zamani, M., Ashtary-Larky, D., Nosratabadi, S., Bagheri, R., Wong, A., Rafiei, M. M., Mahdavi Asiabar, M., Khalili, P., Asbaghi, O., & Davoodi, S. H. (2023). The effects of Gymnema sylvestre supplementation on lipid profile, glycemic control, blood pressure, and anthropometric indices in adults: A systematic review and meta-analysis. Phytotherapy Research, 37(3), 949–964. https://doi.org/10.1002/ptr.7585