Natural Resources for Human Health
eISSN: 2583-1194

Review Article

**Received** 12 May 2026
**Revised** 10 June 2026
**Accepted** 30 June 2026

Natural Resources for Human Health 2026, Vol. 6 (4s): 75–92

**KEYWORDS**:
artificial intelligence; machine learning; natural products; type 2 diabetes mellitus; ethnopharmacology; network pharmacology; QSAR; α-glucosidase

# Artificial Intelligence in Natural Product Antidiabetic Drug Discovery: From Ethnopharmacological Knowledge to Machine Learning-Guided Lead Identification

Dr. Jagannath Nalavade[1], Dr. Rajani Sajjan[2], Shweta Lilhare[3], Dr.Vishnu Suryawanshi[4], Dr. Amar Buchade[5], Dr. Nandkishor P Karlekar[6]

[1]Computer Science and Engineering, MIT School of Computing, MIT Art Design and Technology University, Pune 412201, India.
Email: jen20074u@gmail.com, Orcid ID: 0000-0002-1645-0095
[2]Computer Science and Engineering, MIT School of Computing, MIT Art Design and Technology University, Pune 412201, India.
Email: rajani.sajjan@mituniversity.edu.in, Orcid ID: 0000-0002-2679-5206
[3]AIML, G.H Raisoni College of Engineering, Pune 412207, India
Email id: shweta1641991@gmail.com, Orcid Id:0009-0006-3405-4228
[4]Electrical and Electronics Engineering, MIT School of Computing, MIT Art, Design and Technology University, Pune 412308, India
Email id: vishnusam2007@gmail.com, Orcid Id:0000-0003-0648-2330
[5]Artificial Intelligence and Data Science, BRACT'S Vishwakarma Institute of Technology Pune 411037, India
Email id: amar.buchade@vit.edu, Orcid Id: 0000-0002-6092-6601
[6]Computer Science and Engineering, MIT School of Computing, MIT Art Design and Technology University, Pune 412201, India.
Email id: nkarlekar@gmail.com, Orcid Id: 0009-0009-8046-6452

**ABSTRACT:** When it comes to antidiabetic agents, the core source is plants and the other natural products or their subsequent products in addition to their derived accounts for a sizable proportion of the small molecule drugs that have been allowed from 1981 to 2019.

The traditional method of ethnobotanical examination to a categorised lead is apparently lengthy, and the chemical space of plant secondary metabolites is so huge that it has to be examined comprehensively by test. For minimising the time of the process, machine learning is the solution. The present paper studies the application of these methods throughout the antidiabetic organic product system and calculates the authenticity of the performance as an outcome. The primary test is the molecular target landscape of type two mellitus. Objectives contrast noticeably in their compliance for data-driven modelling, with carbohydrate-digesting enzymes substantially competent served by existing bioactivity data than intracellular signalling proteins.

The pipeline is then drawn through the digitally recorded data into IMPPAT and COCONUT a machine-readable atlas with the selection of molecular representation, traditional and graph-based structures reshaped by AlphaFold and predictive ADMET filtering. The recorded classifier performance throughout this literature collects in the

range of 0.80 and 0.93 in the area under the receiver operating characteristic curve. Many experiments and groups have shown the anticipated hits as outputs including nervonic acid against α-glucosidase and proline-rich peptides against dipeptidyl peptidase-4. There are five methodological flaws that has been observed: tiny and narrow chemically training sets, arbitrary instead of scaffold-based data splitting, resampling applied earlier instead of within cross-validation folds, domain analysis applicably absent, and future experimental testing is limited. Overlooking these factors resulted in accuracy characterises the data partition than the underlying structure activity relationship. The proposed minimal standard is a seven-stage reporting pipeline with the argument that through the machine learning methodology in this field is more precisely described as a candidate trigger than an activity furcating, and the value should be assessed by hit rate in prospective assay than by reviewing classification accuracy. The subsequent research identifies that explainability federated learning across institutional datasets and the integration of metabolomic profiles.

## 1. INTRODUCTION

According to the 11th Edition of IDF Diabetes Atlas the diabetic patents volume may rise to 589 million and projects further rise to 853 million by 2050, with substantial population of these cases will remain undiagnosed till the deadly complexities have developed. (Genitsaridi et al., 2026; Teufel et al., 2026; Duncan et al., 2025). Type 2 diabetic mellitus is going to be the principal chronic disease challenge ahead the world. T2DM will be the major contributor. Glycaemic control in most patients are healed through available oral therapist though there are limitations: gastrointestinal intolerance in the case of α-glucosidase inhibitors, are facing the challenges of sudden weight increase with thiazolidinediones, and the higher cost of treatments and contact with different new incretin-based agents.

Compounds derived from natural plants have shown positive results in these cases. Newman and Cragg have stated that natural products and their derivatives account for a substantial proportion of new small-molecule therapeutics, a pattern sustained across successive editions of their analysis (Newman and Cragg, 2020; Newman, 2022) the survey has the drug approval for four decades. Through two experiments the result is derived- The archetypal α-glucosidase inhibitor, was isolated from an actinomycete, while metformin derives from guanidine chemistry characterised in *Galega officinalis*. The traditional pharmacopoeias of India, China and other regions document several hundred further species employed for conditions consistent with diabetes, a large proportion of which have not been examined using contemporary analytical methods (Ansari et al., 2023).

The major issue found in the experiment is the measuring scale as the selected plant sometime possesses several hundred detectable metabolites. IMPPAT 2.0 accounts approximately 18 thousand phytochemicals across 4 thousand Indian species (Vivek-Ananth et al., 2023), and COCONUT sums a substantially larger collection compiled from more than 50 source databases (Sorokina et al., 2021; Chandrasekhar et al., 2025).

Operating an exhaustive screening of a chemical space of this scale against even a single enzyme is beyond practical measurement. Through Machine learning an effective approach to narrowing this search space is achieved, and the antidiabetic literature of the last 5 years includes several studies that follow a similar design, wherein a classifier is trained using curated bioactivity data, a natural product library is computationally screened, and promising candidates are subsequently selected for experimental validation.

This review significantly observes the body of research instead of merely providing a catalogue of present studies. The discussion is focused on two essential questions. The first states the architecture of the approach: what constitutes a comprehensive artificial intelligence pipeline for the discovery of antidiabetic natural products, and how its individual components are interconnected. The second shows on evaluation: given that the reported accuracies frequently surpass eighty per cent, to what extent are these levels of performance replicated on independent test datasets and supported by experimental validation? This latter question has considerable practical significance, as computational candidate identification is relatively inexpensive compared with experimental verification, and the overall value of the approach ultimately depends on the proportion of computationally nominated candidates that are subsequently validated experimentally.

Significant reviews have already added valuable information to this field. Odugbemi et al. (2024) researched quantitative structure–activity relationship modelling for α-glucosidase, tracing the evolution of descriptor selection and algorithmic approaches across a decade of research. Ansari et al. (2024) reviewed plant-based dietary interventions for diabetes from a pharmacological perspective, while Jiménez-Luna et al. (2020) explored the role of explainability in drug discovery more broadly. Nonetheless, to the best of existing knowledge, no present review systematically traces the complete trajectory from ethnobotanical evidence to experimentally validated lead compounds within the antidiabetic domain, while examining both computational and pharmacological dimensions with comparable rigour. In addition to this, the present literature does not explicitly make differentiations between research outcomes supported by experimental evidence and those that remain primarily computational predictions.

The review is showing the results as follows. Section 2 defines the scope of the review and outlines the criteria used for literature selection. Section 3 develops the discussion across the major stages of the discovery pipeline, encompassing the therapeutic target landscape, digitisation of traditional knowledge, molecular representation, classical and deep learning approaches, network pharmacology, structure-based methods, and predictive toxicology. The section then examines the role of explainability, identifies recurring methodological limitations in the existing literature, and proposes a reporting framework designed to address these shortcomings. Section 4 presents the concluding observations and summarises the key implications of the review.

## 2. SCOPE AND SELECTION OF LITERATURE

### 2.1 Search Strategy

This study follows a narrative-critical review approach instead of a systematic review methodology. Present literature was collected from sources like PubMed, Scopus, Web of Science, and Google Scholar using combinations of search terms spanning three principal domains: the computational domain (artificial intelligence, machine learning, deep learning, QSAR, network pharmacology, virtual screening, and graph neural networks); the therapeutic domain (type 2 diabetes, antidiabetic agents, α-glucosidase, α-amylase, DPP-4, PTP1B, and PPARγ); and the source-material domain (natural products, phytochemicals, medicinal plants, ethnopharmacology, and traditional medicine). The reference lists of the retrieved articles were also examined to identify additional relevant sources, while the citation networks of key methodological studies were reviewed to capture recent applications and emerging developments in the field.

### 2.2 Inclusion and Exclusion Criteria.

Priority is given to the data and studies published from 2020 onward, as the methodological landscape of the field underwent substantial development around this period. Earlier studies were retained when they introduced foundational resources, concepts, or methodologies that remain actively relevant. Studies were considered eligible when they employed computational learning approaches to identify or characterise natural compounds targeting a validated antidiabetic target. Purely synthetic-chemistry-based QSAR studies were excluded unless their methodologies had direct relevance to natural product research. Clinical trials involving herbal preparations without a computational component were also excluded, as were conference abstracts that lacked sufficient methodological detail for meaningful assessment. Particular emphasis was placed on studies reporting experimental validation of computational predictions, as such evidence provides the most direct basis for addressing the central evaluative question of this review.

### 2.3 Synthesis

The performance metrics reported throughout this review are those provided by the original authors and have not been independently recalculated or standardised. This constitutes an important methodological limitation. Performance measures derived from different datasets and evaluated under varying data-splitting strategies are not directly comparable, and the numerical clustering discussed in Section 3.4 should therefore be interpreted with this qualification in mind. Where studies reported the performance of multiple models, the results for the best-performing model are presented, consistent with common reporting practices in the field. However, this approach may introduce a modest upward bias in the overall assessment of reported model performance.

## 3. RESULTS AND DISCUSSION

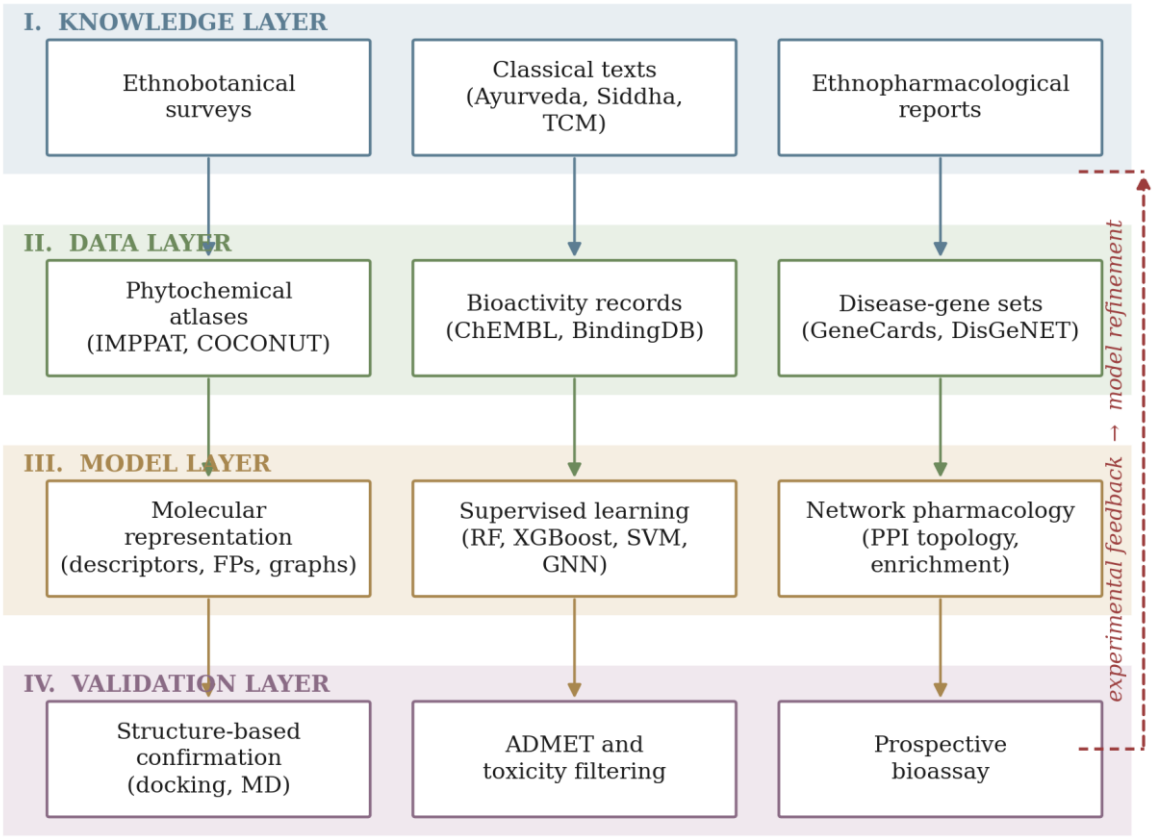


**Figure 1.** Layered workflow linking ethnopharmacological knowledge to experimentally validated antidiabetic leads. Each layer transforms the output of the layer above into a form the next can consume; the dashed return path indicates the experimental feedback that distinguishes a discovery pipeline from a prediction exercise.

### 3.1 The Antidiabetic Target Landscape and Its Consequences for Modelling

T2DM is a multifactorial disorder rather than a disease driven by a single mechanism, and its therapeutic targets can be broadly classified into distinct families that exhibit different characteristics when subjected to computational analysis (Figure 2). This distinction has received comparatively limited attention, despite its importance in determining both the availability and scale of training data and the pharmacological constraints that candidate compounds must satisfy.

Carbohydrate-digesting enzymes constitute the first major target family. Pancreatic α-amylase hydrolyses starch into oligosaccharides, while intestinal α-glucosidase completes their breakdown into absorbable monosaccharides; consequently, inhibition of either enzyme can attenuate the postprandial rise in blood glucose (Odugbemi et al., 2024). Two characteristics make this family particularly well suited to natural product research. First, inhibition occurs within the intestinal lumen, meaning that systemic absorption of the inhibitor is not essential. This reduces the permeability constraints that might otherwise exclude many structurally large polyphenols and glycosides. Second, sustained research and screening efforts have generated a relatively substantial body of bioactivity data, providing the data foundation required for effective supervised machine-learning models.

Intracellular signalling proteins constitute the second major target family and include protein tyrosine phosphatase 1B (PTP1B), AKT1, glycogen synthase kinase-3β (GSK-3β), and insulin receptor substrates. The pharmacological rationale for targeting these proteins is comparatively stronger, as they are more closely associated with the molecular mechanisms underlying insulin resistance. However, their intracellular localisation introduces greater practical challenges. Candidate compounds must first traverse the intestinal epithelium and subsequently gain access to the intracellular environment. Moreover, the available bioactivity datasets are relatively sparse and are predominantly enriched in synthetic chemotypes. Consequently, models trained

on such datasets may have limited generalisability and may perform poorly when extrapolated to structurally distinct plant-derived metabolites.

The incretin axis, with dipeptidyl peptidase-4 (DPP-4) as its principal target, constitutes a third major family with distinct characteristics. A substantial proportion of natural product research in this area focuses on peptides rather than small molecules, fundamentally altering the representation problem: sequence-based descriptors become more appropriate, while the relevant chemical space is combinatorial rather than primarily scaffold-based (Zhang et al., 2024; Lu et al., 2024). A fourth family encompasses transcriptional and inflammatory targets, including PPARγ, AMPK, TNF, and the receptor for advanced glycation end products (RAGE). These targets frequently emerge as hub proteins in network pharmacology analyses involving taxonomically diverse plant species. However, this pattern warrants cautious interpretation, as extensively studied proteins tend to be more comprehensively represented and annotated in interaction databases and are consequently more likely to appear as network hubs, irrespective of their actual biological centrality.

**Table 1.** Principal antidiabetic targets and their tractability for machine learning approaches.

| Target family | Representative targets | Site of action | Data availability | Principal modelling constraint |
|---|---|---|---|---|
| Carbohydrate digestion | α-Amylase, α-glucosidase, MGAM, SGLT-2 | Intestinal lumen / brush border | Comparatively deep | Assay heterogeneity (yeast vs mammalian enzyme) |
| Insulin signalling | PTP1B, AKT1, GSK-3β, IRS1, INSR | Intracellular | Moderate; synthetic-dominated | Permeability; extrapolation from synthetic chemotypes |
| Incretin axis | DPP-4, GLP-1R, GCGR | Plasma membrane / circulation | Moderate; peptide-rich | Representation choice for peptides |
| Transcription and inflammation | PPARγ, AMPK, TNF, IL-6, RAGE | Nuclear / cytoplasmic | Uneven | Annotation bias inflates apparent hub status |

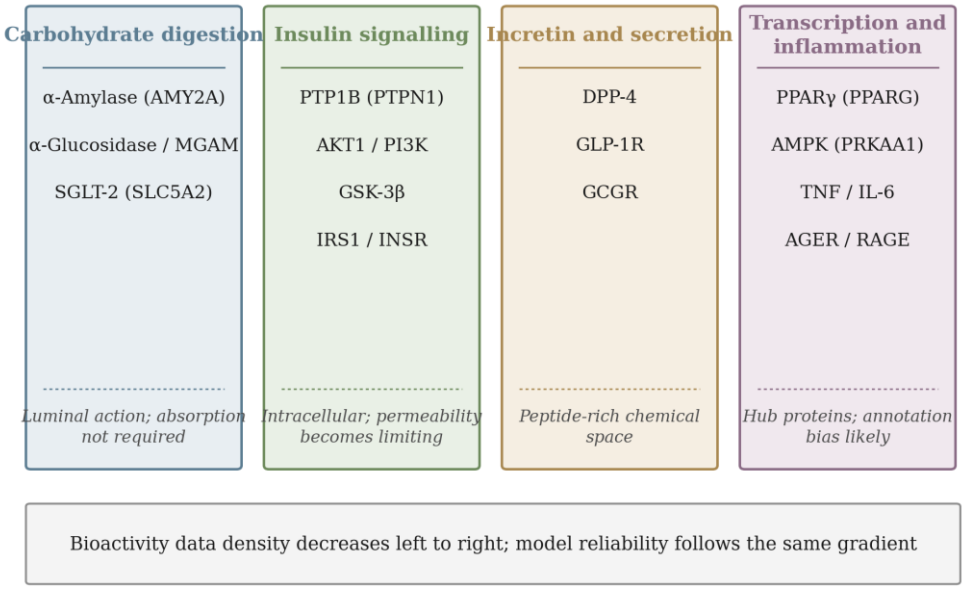


**Figure 2.** Principal antidiabetic target families grouped by mechanism, with the practical consequences of each grouping for data-driven modelling. Bioactivity data density, and therefore model reliability, decreases from left to right.

### 3.2 Digitising Ethnopharmacological Knowledge

The primary stage in any pipeline of this type involves transforming available data and knowledge into a technologically suitable for statistical analysis, a process that is considerably more complex than it might initially appear. Ethnobotanical knowledge is typically represented as plant–indication associations documented in classical texts, ethnographic records, and field surveys, whereas computational models generally require structure–activity relationships linking specific chemical entities to measurable biological effects. Bridging this conceptual and informational gap requires two critical forms of evidence: establishing which bioactive compounds are present in a particular plant or plant part, a question primarily addressed through analytical chemistry, and determining which molecular targets these compounds interact with, a question that falls within the domain of pharmacology.

Several established resources now address the first component of this challenge. IMPPAT 2.0 curates associations between phytochemicals and therapeutic uses across approximately four thousand Indian medicinal plants, with information recorded at the level of individual plant parts. This level of granularity is particularly valuable because the concentrations and profiles of flavonoids and phenolic acids can vary substantially among the leaves, flowers, and roots of the same species (Vivek-Ananth et al., 2023). COCONUT, meanwhile, consolidates openly accessible natural-product datasets into a unified searchable resource containing molecular structures, source organisms, and literature provenance (Sorokina et al., 2021; Chandrasekhar et al., 2025). ChEMBL provides complementary bioactivity data, and recent releases have incorporated natural-product annotations and a natural-product-likeness score, thereby facilitating more efficient extraction of compounds according to their biological and natural-product characteristics (Zdrazil et al., 2024).

No existing resource currently provides a sufficiently reliable mapping between traditional therapeutic indications and specific molecular targets, representing one of the most fundamental data gaps in the field. Classical medical texts may document, for example, the use of a decoction prepared from a particular plant species to treat excessive urination, yet such records rarely indicate whether the observed therapeutic effect involves delayed carbohydrate absorption, enhanced insulin sensitivity, or an entirely different biological mechanism. In practice, this missing link is often bridged through inference or assumption, but the basis for such assumptions is seldom made explicit. Emerging applications of language models to the curation and structuring of traditional medicine corpora offer a potentially valuable route for addressing this gap. However, a critical challenge remains establishing reliable evaluation and validation procedures to verify whether the associations extracted by these models accurately reflect the underlying traditional and pharmacological evidence.

**Table 2.** Principal resources supporting artificial intelligence workflows in natural product antidiabetic discovery.

| Resource | Type | Role in the pipeline |
|---|---|---|
| IMPPAT 2.0 | Phytochemical atlas | Plant-part-resolved constituent lists for Indian medicinal species |
| COCONUT | Natural product collection | Aggregated structures, source organisms and provenance |
| ChEMBL | Bioactivity database | Training data; natural product flags and NP-likeness score |
| SwissTargetPrediction | Target prediction | Ligand-similarity mapping of compounds to human proteins |
| GeneCards / DisGeNET / OMIM | Disease-gene resources | Assembly of disease-associated target sets |
| STRING | Interaction database | Protein–protein network construction and scoring |
| Cytoscape | Network analysis | Topological ranking and hub identification |
| AutoDock Vina | Docking engine | Structure-based confirmation of predicted binding |
| AlphaFold DB | Structure repository | Predicted structures for targets lacking crystallographic data |
| SwissADME / pkCSM / ProTox | ADMET prediction | Pharmacokinetic and toxicity filtering |
| RDKit / scikit-learn / SHAP | Software toolchain | Descriptor generation, model training, feature attribution |

### 3.3 Molecular Representation

A molecule is represented to an algorithm fundamentally constrains what the model can learn, and the choice of representation accounts for a substantial proportion of the variation observed in reported predictive performance. Four broad representation families are commonly used. Physicochemical descriptors encode calculated properties such as molecular weight, lipophilicity, polar surface area, and hydrogen-bonding capacity. These representations are compact and readily interpretable, making them particularly suitable for small datasets, but they inevitably discard detailed structural information. Circular fingerprints, particularly Morgan or extended-connectivity fingerprints (ECFPs), represent molecular structure through the presence or absence of local substructural environments encoded as bit vectors and generally perform well in a range of classification tasks. Substructure keys, such as MACCS keys, represent molecules using a predefined dictionary of chemically meaningful structural patterns, offering a standardized and interpretable representation but with limited coverage of patterns outside the predefined key set. In contrast, molecular graph representations encode atoms as nodes and chemical bonds as edges, allowing graph-based models to learn task-relevant structural representations directly from molecular topology. The choice among these representations therefore involves a trade-off between interpretability, structural information, data requirements, and the model's capacity to learn chemically meaningful features

The proof regarding which molecular representation performs best is less conclusive than might be expected. Jamir et al. (2024), for example, evaluated 35 model–fingerprint combinations for α-glucosidase inhibition and identified the combination of RDKit fingerprints and random forest as the best-performing approach. In contrast, studies involving related biological targets have reported superior performance with circular fingerprints or two-dimensional molecular descriptors. These divergent findings suggest that no single molecular representation consistently outperforms the others across datasets and prediction tasks. Rather, performance appears to depend on an interaction among representation type, dataset size, molecular diversity, target characteristics, and the learning algorithm employed. Accordingly, studies that evaluate only a single molecular representation provide limited evidence about the robustness of their predictive results. A more rigorous benchmarking strategy should therefore compare multiple complementary representations—ideally at least three—under the same data-splitting, preprocessing, and evaluation protocols. Such comparisons can help determine whether observed performance reflects genuine chemical signal or is instead an artefact of a particular representation–model combination.

Plants as natural products further complicate this landscape because they occupy a chemical space that differs substantially from that of conventional synthetic libraries. They are often characterised by greater stereochemical complexity, a higher prevalence of fused and bridged ring systems, greater oxygenation, and higher average molecular weights. Consequently, models trained predominantly on synthetic-compound bioactivity data may be required to extrapolate when applied to plant-derived metabolites, even when the training dataset is large. This issue is particularly important because dataset size alone does not guarantee adequate chemical-space coverage. Applicability-domain (AD) analysis provides a means of identifying compounds that fall outside, or at the margins of, the chemical space represented in the training data. Its omission is therefore especially consequential in natural-product QSAR studies, where structural novelty and distributional differences between training and application domains can substantially affect predictive reliability. Reporting AD measures alongside conventional performance metrics is thus essential for assessing whether a model's predictions for plant metabolites represent genuine interpolation within learned chemical space or potentially unreliable extrapolation beyond it.

### 3.4 Classical Machine Learning for Antidiabetic Bioactivity Prediction

A substantial body of published research in this area has employed conventional supervised learning algorithms, and the overall pattern of findings is sufficiently consistent to permit several broad conclusions. Ensemble tree-based methods, particularly random forest and gradient-boosting variants, frequently emerge among the best-performing models. Support vector machines also remain competitive, particularly when combined with appropriate molecular representations and feature-selection strategies. By contrast, neural-network approaches have generally offered limited or inconsistent advantages at the relatively small dataset sizes typical of natural-product bioactivity studies. These findings suggest that model complexity does not necessarily translate into superior predictive performance when data are limited and chemically heterogeneous. Representative studies illustrating these trends are summarised in Table 3.

Jamir et al. (2024) offers one of the most inclusive examples of this methodology. Using a dataset of 537 reported plant secondary metabolites with α-glucosidase inhibitory activity, the authors evaluated 35 model–representation combinations and

identified random forest as the best-performing model, achieving an accuracy of 83.7% and an area under the receiver operating characteristic curve (AUC–ROC) of 0.803. The selected model was subsequently applied to an in-house library of 5,810 metabolites, from which 24 predicted candidates were selected for experimental validation. Thirteen of these compounds demonstrated α-glucosidase inhibitory activity, including seven with $IC_{50}$ values lower than that of the reference inhibitor acarbose; nervonic acid emerged as the most promising candidate. The experimental validation rate—13 confirmed hits among 24 tested compounds, or approximately 54%—is particularly informative when assessing the practical utility of the model. Unlike retrospective accuracy or AUC estimates obtained from the training and validation datasets, this hit rate reflects the model's ability to identify biologically active compounds from previously untested chemical space. It therefore provides a more direct indication of the translational value of machine-learning-assisted virtual screening.

Research on dipeptidyl peptidase-4 (DPP-4) follows a broadly similar pattern. Lu et al. (2024) developed a quantitative structure–activity relationship (QSAR) model using seven descriptors selected through a genetic algorithm. The model achieved a test-set coefficient of determination ($R^2$) of 0.816 and was subsequently used to screen 1,668 natural flavonoid derivatives, from which three candidates were advanced to molecular-dynamics simulations for further validation. Zhang et al. (2024) evaluated five gradient-boosting algorithms for predicting DPP-4-inhibitory peptides and reported an area under the receiver operating characteristic curve (AUC–ROC) of 0.92 for Light GBM. The model was then applied to virtual proteolysis of bovine milk proteins, leading to the identification of candidate peptides, two of which were experimentally confirmed. Earlier, Hermansyah et al. (2021) combined QSAR-based classification with molecular docking for DPP-4 inhibition, establishing a methodological framework subsequently adopted by several studies targeting the same enzyme. Srisongkram et al. (2022) adopted a complementary approach by screening 31 Thai plant extracts against both α-glucosidase and DPP-4 and applying classification analysis to compare their predictability. Their findings indicated that α-glucosidase inhibitory activity was more readily predicted than DPP-4 activity. Collectively, these studies demonstrate a recurring workflow in natural-product research: computational model development, virtual screening or prioritisation, structural validation, and, where feasible, experimental confirmation.

Two observations emerge from this body of evidence. First, reported predictive performance tends to cluster within a relatively narrow range, approximately 0.80–0.93 in AUC, across different targets, algorithms, datasets, and research groups. Such consistency may reflect less a genuine equivalence among modelling approaches than a performance ceiling imposed by the quality, heterogeneity, and measurement consistency of the underlying bioactivity data. $IC_{50}$ values aggregated across laboratories, assay platforms, experimental protocols, and enzyme sources are subject to substantial inter-assay and measurement variability, which no machine-learning algorithm can fully eliminate. Second, the relatively small proportion of studies that proceed to prospective experimental validation generally report encouraging but substantially lower hit rates than might be anticipated from their retrospective accuracy or AUC values. This discrepancy is consistent with a shift from interpolation within the training chemical space to prediction in less familiar or novel chemical space, where applicability-domain limitations become more consequential. Prospective validation should therefore be treated as a critical complement to conventional statistical metrics when assessing the real-world utility and generalisability of computational models for natural-product discovery.

**Table 3.** Representative machine learning studies in antidiabetic natural product discovery. Performance figures are as reported by the original authors.

| Study | Target | Approach | Reported performance | Experimental follow-up |
|---|---|---|---|---|
| Jamir et al. (2024) | α-Glucosidase | 35 models × 7 fingerprints; RF with RDKit best | ACC 83.7%; ROC-AUC 0.803 | 24 compounds assayed; 13 active; nervonic acid identified |
| Srisongkram et al. (2022) | α-Amylase, α-glucosidase | Classification and multivariate analysis on 31 plant extracts | α-Glucosidase activity more | In vitro screening of all 31 extracts |

| Study | Target | Approach | Reported performance | Experimental follow-up |
|---|---|---|---|---|
| | | | predictable than α-amylase | |
| Lu et al. (2024) | DPP-IV | GA-selected QSAR on 7 descriptors; 1668 flavonoids screened | $R^2$ (test) 0.816; MAE 0.14 | 3 leads carried to 100 ns MD and MM/PB(GB)SA |
| Zhang et al. (2024) | DPP-4 (peptides) | GBDT, XGBoost, LightGBM, CatBoost, RF; virtual proteolysis | ROC-AUC 0.92 ± 0.01 (LightGBM) | 2 milk-derived peptides confirmed active |
| Khaldan et al. (2024) | α-Glucosidase | 3D-QSAR with ADMET, docking, MD and quantum studies | Validated 3D-QSAR models on 31 triazole derivatives | Computational only |
| Asnawi et al. (2024) | α-Glucosidase | Pharmacophore modelling and virtual screening of curculigoside A derivatives | Docking and MD-based ranking | Computational only |
| Ononamadu et al. (2025) | Multiple (T2DM) | Network pharmacology, docking and MD for Melissa officinalis | 134-node, 398-edge network; EGFR, SRC, AKT1 as hubs | Computational only |
| Ononamadu and Seidel (2024) | Multiple (T2DM) | Network pharmacology with ADME and drug-likeness for Salvia officinalis | Hub and pathway identification | Computational only |

### 3.5 Deep Learning and Graph-Based Approaches

Graph neural networks (GNNs) symbolises molecules as graphs in which atoms constitute nodes and chemical bonds constitute edges, enabling the model to learn molecular representations through iterative message passing among neighbouring atoms. This approach reduces the need for researchers to specify a descriptor set or manually engineer structural features in advance. On large benchmark datasets, GNNs have frequently demonstrated superior performance compared with conventional descriptor-based approaches. However, this advantage is considerably less consistent in natural-product antidiabetic research, where datasets typically contain only several hundred to a few thousand compounds. Under such data-constrained conditions, classical ensemble methods can often match or even outperform GNNs. This apparent discrepancy is better understood because of the relationship between model complexity and available training data than as a limitation of the GNN architecture itself. With insufficient samples to adequately constrain a high-capacity model, GNNs may be more susceptible to overfitting and may fail to realise their representational advantage. Their effectiveness is therefore likely to depend strongly on dataset size, chemical diversity, representation quality, and the availability of sufficiently large and chemically relevant training data.

The latest research in which it is observed that two recent developments may alter this balance. Transfer learning and self-supervised pre-training enable models to learn general chemical representations from large collections of unlabelled molecular

data and subsequently adapt those representations to relatively small, labelled datasets. This strategy is particularly well suited to the data-constrained setting characteristic of natural-product research, where experimentally validated bioactivity data are often limited. A related development involves molecular language models trained on SMILES strings and subsequently fine-tuned for specific prediction tasks. Such approaches have begun to emerge in α-glucosidase research, with data-augmentation strategies employed to mitigate the limitations imposed by small training datasets. However, these methods remain less extensively validated than conventional QSAR and ensemble-learning approaches within this specific research domain. Their apparent gains should therefore be interpreted cautiously, particularly where external validation, chemical-space diversity, and applicability-domain analysis are limited. The reporting principles discussed in Section 3.10—including transparent data partitioning, independent validation, comparison with appropriate baselines, and explicit assessment of generalisability—are consequently at least as important for these higher-capacity approaches as they are for classical machine-learning models.

Generative models represent a more prospective direction for natural-product discovery. Architectures capable of generating novel molecular structures conditioned on desired biological or physicochemical properties have advanced rapidly, while natural-product scaffolds provide an attractive starting point because their structural diversity and evolutionary selection may favour biologically relevant chemical features. The principal challenge, however, lies in synthetic accessibility. A structure generated from a complex terpenoid or other structurally elaborate natural-product scaffold may be no easier to synthesise, isolate, or modify than the parent compound itself, substantially limiting its practical value. Consequently, generative approaches in this domain will need to be integrated with retrosynthetic analysis, synthetic-accessibility scoring, and realistic assessment of precursor availability. Their utility should ultimately be judged not simply by the novelty or predicted activity of the generated structures, but by whether those structures can be synthesised or otherwise obtained and subsequently validated experimentally. Such integration will be essential for translating generative molecular design from computational novelty generation into practically actionable natural-product discovery.

**3.6 Network Pharmacology as a Systems-Level Complement**

Where quantitative structure–activity relationship (QSAR) modelling asks whether a compound is likely to inhibit a particular protein, network pharmacology addresses a broader question: which molecular targets are collectively engaged by the multiple constituents of a medicinal plant, and how are those targets functionally and structurally interconnected? The two approaches are therefore complementary rather than competing, and their integration has become a common computational framework for investigating the multi-component and multi-target nature of medicinal plants.

A conventional network-pharmacology workflow begins by predicting potential protein targets for individual phytoconstituents, often using SwissTargetPrediction (Daina et al., 2019). These predicted targets are then intersected with disease-associated genes obtained from resources such as GeneCards, DisGeNET, and OMIM (Piñero et al., 2020). The overlapping targets are subsequently mapped onto a protein–protein interaction (PPI) network, commonly using STRING (Szklarczyk et al., 2023, 2025), and visualised and analysed in Cytoscape, where network-topological measures such as degree, betweenness, and closeness centrality are used to identify highly connected candidate hubs. Finally, functional enrichment analyses, typically based on Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, are used to interpret the biological processes and signalling pathways associated with the predicted targets.

Applied to *Melissa officinalis*, this workflow produced a network comprising 134 nodes and 398 edges, with EGFR, SRC, AKT1, TNF, and ESR1 emerging among the principal hub targets (Ononamadu et al., 2025). Comparable network-level patterns have been reported for *Salvia officinalis* and fenugreek, further illustrating how network pharmacology can provide a systems-level interpretation of the potentially interconnected actions of plant-derived compounds (Ononamadu & Seidel, 2024; Luo et al., 2023). However, such networks should be interpreted as hypothesis-generating models rather than direct evidence of biological interaction, because predicted targets, database-derived disease associations, and network centrality do not by themselves establish pharmacological activity or causal relevance.

This convergence warrants closer scrutiny. One possible interpretation is that taxonomically unrelated medicinal plants genuinely converge on central nodes within insulin-signalling and related metabolic pathways. An alternative explanation, however, is that the databases underlying these analyses are disproportionately populated by well-characterised proteins and pathways. Under such conditions, highly studied targets such as AKT1 and PPARγ may repeatedly emerge as network hubs largely independently of the specific set of input compounds. Both mechanisms are likely to contribute to some extent, but

distinguishing genuine biological convergence from database-driven bias requires either experimental validation or an appropriately constructed null model based on randomly selected compounds or targets. Few published network-pharmacology studies incorporate either approach.

A further limitation is that network pharmacology generally prioritises proteins according to network topology rather than experimentally measured binding affinity or functional activity. A highly connected node may therefore be biologically important within the constructed network without necessarily being a direct molecular target of any phytoconstituent. This distinction is crucial when interpreting computational predictions. Machine-learning-based QSAR and target-prediction models address a complementary aspect of the problem by estimating compound–target relationships from molecular features and experimentally observed activity data. Network pharmacology consequently provides a systems-level view of potential target interactions, whereas machine learning can provide compound-level predictions of biological activity. Their integration is therefore potentially more informative than either approach in isolation, if database bias, prediction uncertainty, and applicability-domain limitations are explicitly assessed.

**3.7 Structure-Based Methods and the Effect of AlphaFold**

Molecular docking remains the principal structure-based virtual-screening approach in this literature, with AutoDock Vina among the most widely used platforms (Eberhardt et al., 2021). Although docking provides a computationally efficient means of exploring potential ligand–protein binding modes, its scoring functions generally show only a modest correlation with experimentally measured binding affinities. This limitation is well recognised, yet it is often overlooked when interpreting docking results. Differences in predicted binding energies of only a few tenths of a kilocalorie per mole are sometimes treated as evidence of meaningful differences in ligand ranking, despite the uncertainty inherent in the scoring function and the underlying structural model. Machine-learning-based scoring functions, trained on experimentally characterised protein–ligand complexes, offer a potential improvement by learning interaction patterns that conventional empirical scoring functions may not capture. However, these approaches introduce their own methodological risks, particularly when training and test complexes are structurally similar. Without rigorous scaffold-level or protein-level data splitting and truly independent validation, a model may memorise characteristics of its training complexes rather than learn transferable principles of molecular recognition. Consequently, machine-learning scoring functions should be evaluated not only by predictive accuracy but also by their ability to generalise across chemically and structurally distinct protein–ligand systems.AlphaFold has changed the availability side of the problem substantially (Jumper et al., 2021), and the associated structure database has extended coverage across proteomes that were previously inaccessible (Varadi et al., 2022). AlphaFold 3 extends prediction to biomolecular complexes including protein–ligand interactions (Abramson et al., 2024). For antidiabetic work the practical benefit is uneven, since the principal targets already have good experimental structures; the benefit is greater for secondary or less-studied proteins emerging from network analysis. A caution applies regardless: predicted structures represent a single conformational state and may not reproduce the induced-fit behaviour of a ligand-bound active site, so docking into an unvalidated predicted structure is a weaker inference than docking into a co-crystal.

Molecular docking remains the principal structure-based virtual-screening approach in this literature, with AutoDock Vina among the most widely used platforms (Eberhardt et al., 2021). Although docking provides a computationally efficient means of exploring potential ligand–protein binding modes, its scoring functions generally show only a modest correlation with experimentally measured binding affinities. This limitation is well recognised, yet it is often overlooked when interpreting docking results. In particular, differences in predicted binding energies of only a few tenths of a kilocalorie per mole are sometimes treated as evidence of meaningful differences in ligand ranking, despite the uncertainty inherent in the scoring function and the underlying structural model. Machine-learning-based scoring functions, trained on experimentally characterised protein–ligand complexes, offer a potential improvement by learning interaction patterns that conventional empirical scoring functions may not capture. However, these approaches introduce their own methodological risks, particularly when training and test complexes are structurally similar. Without rigorous scaffold-level or protein-level data splitting and truly independent validation, a model may memorise characteristics of its training complexes rather than learn transferable principles of molecular recognition. Consequently, machine-learning scoring functions should be evaluated not only by predictive accuracy but also by their ability to generalise across chemically and structurally distinct protein–ligand systems.3.8 Predictive ADMET and Toxicity

Natural origin should not be interpreted as evidence of safety; consequently, predictive toxicology warrants greater attention in the computational study of medicinal plants than it has traditionally received. Tools such as SwissADME, pkCSM, and ProTox can estimate absorption, distribution, metabolism, excretion (ADME), and a range of toxicity endpoints directly from molecular structure, at sufficiently low computational cost to permit screening across large compound libraries.

Two important qualifications, however, should accompany such predictions. First, the training datasets underlying many of these platforms are dominated by synthetic, drug-like compounds. Their predictive reliability may therefore be reduced for chemical classes that are particularly prominent in natural products, including glycosides, saponins, tannins, and highly hydroxylated polyphenols. Predictions for these compounds should consequently be treated as estimates subject to potentially substantial uncertainty rather than as definitive safety assessments. Second, oral-absorption criteria must be interpreted in relation to the intended pharmacological mechanism. A compound designed to inhibit intestinal α-glucosidase, for example, acts locally at the intestinal brush border and does not necessarily need to undergo systemic absorption to produce a therapeutic effect. Excluding such a compound solely because it fails conventional Lipinski-style drug-likeness or predicted-bioavailability thresholds could therefore eliminate potentially valuable candidates on criteria that are mechanistically inappropriate for the target.

A further underexplored application is herb–drug interaction prediction, particularly the assessment of potential inhibition or induction of cytochrome P450 enzymes and other drug-metabolising systems. This issue has clear clinical relevance because herbal preparations are frequently consumed alongside conventional antidiabetic medications. Integrating ADME, toxicity, and interaction prediction into natural-product screening would therefore provide a more realistic assessment of therapeutic potential and translational risk than activity prediction alone.

### 3.9 Explainability

A model that ranks compounds without indicating the basis for its predictions has limited utility for medicinal chemists, making model explainability an increasingly central concern in computational drug discovery (Jiménez-Luna et al., 2020; Alizadehsani et al., 2024; Lavecchia, 2025). Feature-attribution methods, particularly SHAP (SHapley Additive exPlanations), are now widely used to identify the molecular descriptors or structural features that contribute most strongly to individual predictions. The practical value of such explanations lies in determining whether the model recovers structure–activity relationships that are independently supported by experimental evidence. For example, if a model trained solely on bioactivity data identifies hydrogen-bond donor count and aromatic ring count as important predictors of α-glucosidase inhibition, and structural studies independently demonstrate that ligand recognition within the enzyme's active site depends substantially on hydrogen-bonding interactions with catalytic residues, the convergence provides indirect evidence that the model has captured chemically meaningful relationships rather than dataset-specific artefacts.

The limitations of explainability methods, however, require careful consideration. Post hoc attribution methods explain the behaviour of a predictive model, not necessarily the underlying molecular mechanism, and these two levels of explanation may diverge. Furthermore, attribution scores can be sensitive to the choice of explanation method, model architecture, feature representation, and input perturbations, potentially producing unstable interpretations even for closely related molecules. Explainability should therefore be regarded primarily as a diagnostic and hypothesis-generating tool. It can reveal whether a model relies on chemically plausible features, identify potentially spurious correlations, and guide subsequent experimental investigation, but it should not by itself be interpreted as direct mechanistic evidence. Robust mechanistic conclusions require convergence among model explanations, established chemical knowledge, structural biology, and experimental validation.

### 3.10 Recurring Methodological Weaknesses

Five methodological problems recur across this literature with sufficient frequency to be regarded as systemic rather than incidental. They are summarised in Table 4 alongside recommended practices for addressing them.

First, training datasets are often small and chemically narrow. Datasets comprising only a few hundred compounds and drawn from a limited number of scaffold families provide a weak basis for claims of broad generalisability. Odugbemi et al. (2024) identified limited dataset size and diversity as major constraints on generalisation in α-glucosidase modelling specifically. Increasing the number of compounds alone is not sufficient; adequate coverage of relevant chemical space is equally important.

Second, random data splitting remains the predominant validation strategy. In natural-product libraries containing numerous structurally related analogues, random partitioning can place near-identical compounds in both training and test sets, thereby inflating apparent predictive performance through information leakage. Scaffold-based splitting, which separates compounds according to core structural frameworks, provides a more stringent assessment of generalisation to chemically distinct molecules and should be preferred where dataset size permits.

Third, class imbalance is frequently addressed using synthetic oversampling before data partitioning. Applying oversampling to the complete dataset allows information from compounds subsequently assigned to the test set to influence the construction of synthetic training examples, resulting in optimistic performance estimates. Resampling should instead be performed within each training fold after the train–test or cross-validation split, thereby preventing information leakage and preserving the independence of the evaluation data.

Fourth, the applicability domain (AD) is rarely defined or reported adequately. Without an explicit AD, a model can generate a seemingly confident prediction for a molecule that is structurally unlike anything represented in its training data, while providing no indication that the prediction constitutes extrapolation. Given the substantial differences between natural-product and conventional synthetic chemical space discussed in Section 3.3, this limitation is particularly consequential for natural-product QSAR modelling. AD assessment should therefore accompany predictive performance metrics and identify compounds for which model confidence is insufficient.

Finally, and most importantly, most published studies stop at computational prioritisation rather than prospective experimental validation. High accuracy or AUC on a held-out test set demonstrates performance under a retrospective evaluation framework, but it does not establish that a model can successfully identify genuinely active compounds in previously unexplored chemical space. Prospective validation, in which model-selected candidates are experimentally tested without prior knowledge of their activity, provides substantially stronger evidence of practical utility. The difference is fundamental: retrospective metrics demonstrate that a model can reproduce patterns within an available dataset, whereas a prospective hit rate demonstrates that the complete computational pipeline can generate experimentally actionable discoveries. Consequently, prospective validation should be regarded as the strongest available test of translational performance in machine-learning-assisted natural-product discovery.

**Table 4.** Recurring methodological weaknesses and recommended practice.

| Weakness | Why it matters | Recommended practice |
|---|---|---|
| Small, chemically narrow training sets | Limits generalisation beyond the scaffolds represented | Report scaffold diversity; aggregate across studies; consider federated training |
| Random data splitting | Places close analogues on both sides of the partition, inflating accuracy | Use scaffold-based or time-based splitting and report which was used |
| Resampling before partitioning | Synthetic samples informed by test data give optimistic estimates | Apply SMOTE or equivalent within each cross-validation fold only |
| No applicability domain | Confident predictions returned for compounds unlike anything in training | Define a similarity or leverage threshold; flag out-of-domain predictions |
| No prospective testing | Retrospective accuracy is a much weaker claim than hit rate | Test a defined subset experimentally and report the hit rate obtained |

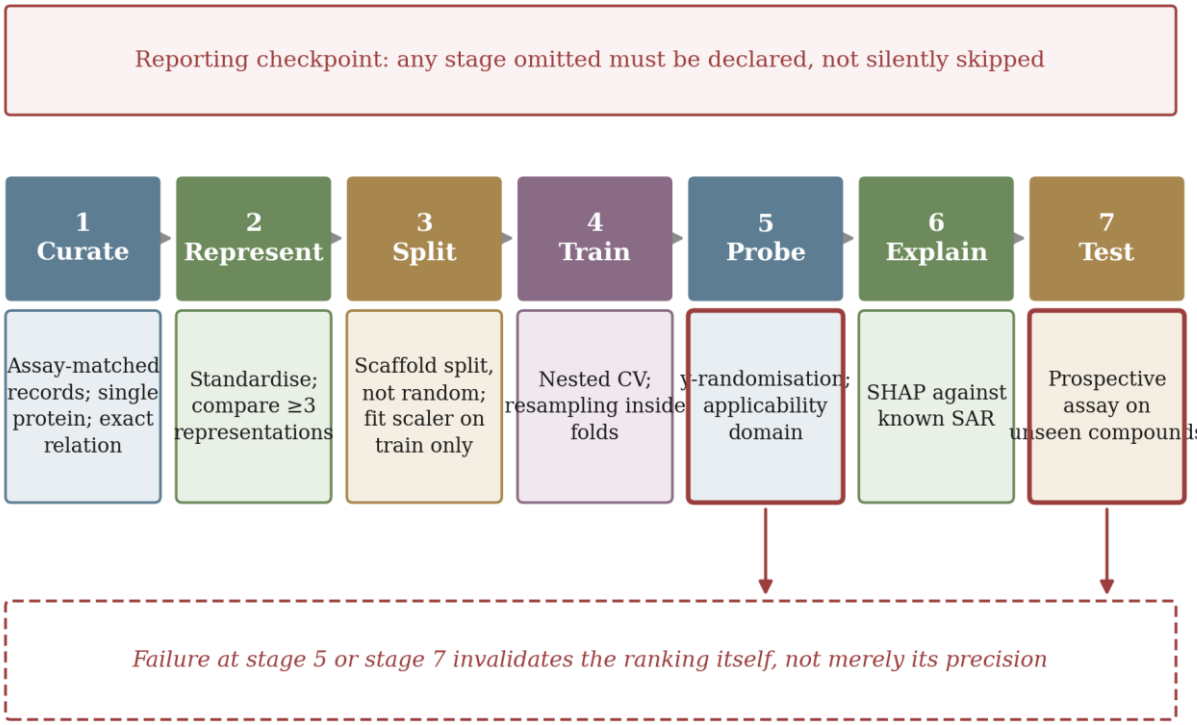


**Figure 3.** Recommended reporting pipeline for bioactivity models built on natural product data. The two checkpoints marked in red distinguish a model whose ranking can be trusted from one whose reported accuracy describes the data partition rather than the underlying chemistry.

### 3.11 A Proposed Reporting Standard

Figure 3 presents a seven-stage pipeline proposed as a minimum methodological standard for bioactivity modelling using natural-product data. The individual components are not novel; all represent established good practices in cheminformatics and machine-learning research. The contribution of the proposed framework is instead to argue that these practices should be treated as essential methodological requirements, with any omission explicitly reported rather than left unacknowledged.

The seven stages comprise: (1) rigorous data curation restricted to assay-matched, single-protein records with clearly defined activity endpoints and exact activity relationships; (2) structural standardisation and comparison of at least three complementary molecular representations; (3) scaffold-based data splitting, with all scaling and preprocessing parameters fitted exclusively on the training partition; (4) nested cross-validation, with any class-balancing or resampling procedures performed strictly within the relevant training folds; (5) y-randomisation testing combined with an explicit definition of the model's applicability domain; (6) feature-attribution analysis evaluated against experimentally established structure–activity relationships; and (7) prospective experimental validation using compounds excluded from all training, model-selection, and validation datasets.

Stages 5 and 7 are deliberately emphasised in Figure 3 because their absence affects the validity of the resulting compound ranking rather than merely its numerical precision. Y-randomisation helps determine whether apparent predictive performance arises from genuine structure–activity relationships rather than chance correlations, while applicability-domain analysis establishes whether predictions represent interpolation within the learned chemical space or unsupported extrapolation. Prospective validation provides the strongest test of whether computational prioritisation translates into experimentally confirmed activity.

Studies that omit these stages may still provide useful exploratory evidence and can contribute to hypothesis generation. However, their conclusions should be framed accordingly: retrospective computational performance should not be presented as equivalent to demonstrated prospective predictive utility.

### 3.12 Future Directions

Four directions appear particularly promising for advancing computational research on medicinal plants and natural-product bioactivity. First, federated learning could enable models to be trained across bioactivity datasets held by multiple institutions without requiring the underlying data to be transferred or centrally pooled. This approach could directly address the persistent problem of limited training data while preserving institutional data governance and privacy. Pharmaceutical consortia have already demonstrated the feasibility of federated approaches for other predictive tasks, suggesting considerable potential for their application to natural-product bioactivity modelling.

Second, integrating metabolomic profiles with experimentally measured activity data could provide a more direct route to attributing the biological activity of crude extracts to specific constituents without requiring complete prior fractionation. Such an approach is particularly relevant to traditional medicinal preparations, where therapeutic effects may arise from complex mixtures rather than isolated compounds. Linking metabolomic signatures to bioactivity could therefore help identify candidate active constituents while preserving information about the chemical context in which they occur.

Third, computational pipelines should increasingly move from modelling individual compounds to compound combinations. Traditional medicinal preparations are mixtures, and their biological effects may involve additive, synergistic, or antagonistic interactions among constituents. A workflow that identifies only single-molecule candidates may consequently overlook an important characteristic of the material being studied. Although synergy prediction is substantially more challenging than single-compound activity prediction and the available training data remain limited, modelling combinations represents a closer match between computational methodology and the pharmacological reality of many traditional preparations.

Finally, the field would benefit from closing the prediction–experiment feedback loop through automated or semi-automated experimental screening. Computationally prioritised compounds could be tested prospectively, with the resulting activity data incorporated into subsequent rounds of model training and candidate selection. Such iterative learning would allow the model to expand its applicability domain through experimentally verified observations rather than merely characterising its existing limitations. In this sense, active-learning workflows that integrate prediction, experimental testing, and model retraining offer a potentially important route from static computational screening towards continuously improving natural-product discovery systems.

## 4. CONCLUSION

Machine learning has established a credible and increasingly useful role in antidiabetic natural-product discovery. Across targets and modelling approaches, reported classifier performance generally falls within an AUC range of approximately 0.80–0.93, while studies that have progressed to experimental validation have confirmed a meaningful proportion of their computationally prioritised candidates, including nervonic acid against α-glucosidase and several proline-rich peptides against dipeptidyl peptidase-4. Relative to untargeted screening across chemical libraries containing tens of thousands of metabolites, such computational prioritisation represents a substantial methodological advance by concentrating experimental resources on a smaller and more tractable set of candidates.

The contribution of machine learning, however, is more appropriately characterised as triage and prioritisation than definitive prediction. A model that reduces a library of 20,000 compounds to a few dozen candidates for experimental testing performs a valuable function even if a substantial proportion of those candidates subsequently prove inactive. By contrast, a model described as reliably predicting biological activity on the basis of retrospective accuracy obtained from a random data split, without an explicitly defined applicability domain or prospective validation, makes a considerably stronger claim than the evidence warrants.

This distinction has important practical and epistemological consequences. It determines whether computational outputs are treated as testable hypotheses requiring experimental confirmation or as established findings upon which subsequent research can safely build. Given the methodological limitations identified in Section 3.10—including small and chemically narrow datasets, random rather than scaffold-based splitting, potential information leakage, inadequate applicability-domain

assessment, and limited prospective validation—the former interpretation remains the more scientifically defensible position for a substantial proportion of the published literature. The field's future progress will therefore depend less on incremental gains in retrospective performance metrics than on rigorous external validation, prospective testing, transparent reporting, and closer integration between computational predictions and experimental evidence.

Three priorities emerge from this review. First, reporting practices should converge towards a methodological standard comparable to that proposed in Section 3.11, with prospective experimental validation treated as a primary measure of a computational pipeline's practical value rather than as an optional extension. Greater emphasis should also be placed on transparent data curation, chemically meaningful validation strategies, applicability-domain assessment, and explicit reporting of negative results.

Second, the persistent problem of data scarcity—which likely contributes to the performance ceiling reflected in the relatively narrow range of reported predictive metrics—needs to be addressed directly. Federated learning, collaborative data-sharing frameworks, and the systematic deposition of both positive and negative bioactivity results could substantially expand the diversity and reliability of training datasets while reducing institutional and publication biases.

Finally, the persistent mismatch between single-compound modelling and the multi-component nature of traditional medicinal preparations warrants considerably greater attention. Traditional formulations rarely consist of isolated molecules; their pharmacological effects may arise from combinations of constituents acting through complementary, additive, or synergistic mechanisms. Computational frameworks capable of modelling these interactions could therefore provide a more faithful representation of traditional medicine than approaches focused exclusively on individual compounds.

Ultimately, extracting pharmacologically meaningful information from the knowledge preserved in classical pharmacopoeias represents a promising application of machine learning. Realising this potential, however, requires the field to match the strength of its methodological standards to the strength of the claims it makes. Machine learning can substantially accelerate the transition from traditional knowledge to testable pharmacological hypotheses, but its contribution will be most valuable when computational predictions remain transparently connected to reproducible data and rigorous experimental validation.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the management of **MIT Art, Design and Technology University** Pune for providing the necessary support and resources to successfully complete this research.

## AUTHOR CONTRIBUTIONS

Author One: Conceptualization, Literature Search, Formal Analysis, Author two: Writing — Original Draft, Visualization. Author Three: Methodology, Literature Screening, Author Four, Five and Six: Writing — Review & Editing, Supervision. All authors have read and approved the final version of the manuscript.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

## ETHICS STATEMENT

This review did not involve human participants, animal subjects or the collection of primary data, and ethical approval was therefore not required.

## DATA AVAILABILITY STATEMENT

No new data were generated for this review. All studies discussed are available from their original publishers, and the search terms applied are given in Section 2.1.

## REFERENCES

[1] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al., 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500.

[2] Alizadehsani, R., Oyelere, S.S., Hussain, S., et al., 2024. Explainable artificial intelligence for drug discovery and development – a comprehensive survey. IEEE Access 12, 35796–35812.
[3] Ansari, P., Samia, J.F., Khan, J.T., Rafi, M.R., Rahman, M.S., Rahman, A.B., Abdel-Wahab, Y.H.A., Seidel, V., 2023. Protective effects of medicinal plant-based foods against diabetes: a review on pharmacology, phytochemistry, and molecular mechanisms. Nutrients 15(14), 3266.
[4] 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.
[5] Alkhatib A. J., (2026). Integrated Computational Retinal Vascular Morphometry Using Fractal, Skeleton-Based, and Phi-Related Features for Diabetic Retinopathy and Glaucoma. International Journal of Artificial Intelligence, Machine Learning and Data Analytics, 1(1), 43-50. https://www.ijaimda.org/index.php/ijaimda/article/view/3
[6] Asnawi, A., Mieldianisa, S., Aligita, W., Yuliantini, A., Febrina, E., 2024. Integrative computational approaches for designing novel alpha-glucosidase inhibitors based on curculigoside A derivatives: virtual screening, molecular docking, and molecular dynamics. Journal of HerbMed Pharmacology 13(2), 308–323.
[7] Chandrasekhar, V., Rajan, K., Kanakam, S.R.S., et al., 2025. COCONUT 2.0: a comprehensive overhaul and curation of the collection of open natural products database. Nucleic Acids Research 53(D1), D634–D643.
[8] Daina, A., Michielin, O., Zoete, V., 2019. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. Nucleic Acids Research 47(W1), W357–W364.
[9] Duncan, B.B., Magliano, D.J., Boyko, E.J., 2025. IDF Diabetes Atlas 11th edition 2025: global prevalence and projections for 2050. Nephrology Dialysis Transplantation 41(1), 7–9.
[10] Eberhardt, J., Santos-Martins, D., Tillack, A.F., Forli, S., 2021. AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. Journal of Chemical Information and Modeling 61(8), 3891–3898.
[11] Genitsaridi, I., Salpea, P., Salim, A., Sajjadi, S.F., Tomic, D., James, S., Thirunavukkarasu, S., Magliano, D.J., Boyko, E.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.
[12] Hermansyah, O., Bustamam, A., Yanuar, A., 2021. Virtual screening of dipeptidyl peptidase-4 inhibitors using quantitative structure-activity relationship-based artificial intelligence and molecular docking of hit compounds. Computational Biology and Chemistry 95, 107597.
[13] Jamir, L., et al., 2024. Employing machine learning models to predict potential α-glucosidase inhibitory plant secondary metabolites targeting type-2 diabetes and their in vitro validation. Journal of Chemical Information and Modeling 64(24), 9150–9162.
[14] Jiménez-Luna, J., Grisoni, F., Schneider, G., 2020. Drug discovery with explainable artificial intelligence. Nature Machine Intelligence 2(10), 573–584.
[15] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.
[16] Khaldan, A., Bouamrane, S., El-mernissi, R., Ouabane, M., Alaqarbeh, M., Maghat, H., Ajana, M.A., Sekkat, C., Bouachrine, M., Lakhlifi, T., Sbai, A., 2024. Design of new α-glucosidase inhibitors through a combination of 3D-QSAR, ADMET screening, molecular docking, molecular dynamics simulations and quantum studies. Arabian Journal of Chemistry 17(3), 105656.
[17] Lavecchia, A., 2025. Explainable artificial intelligence in drug discovery: bridging predictive power and mechanistic insight. WIREs Computational Molecular Science 15(5), e70049.
[18] Lu, G., Pan, F., Li, X., Zhu, Z., Zhao, L., Wu, Y., Tian, W., Peng, W., Liu, J., 2024. Virtual screening strategy for anti-DPP-IV natural flavonoid derivatives based on machine learning. Journal of Biomolecular Structure and Dynamics 42(13), 6645–6659.
[19] Luo, W., Deng, J., Huang, J., et al., 2023. Integration of molecular docking, molecular dynamics and network pharmacology to explore the multi-target pharmacology of fenugreek against diabetes. Journal of Cellular and Molecular Medicine, doi:10.1111/jcmm.17787.
[20] Newman, D.J., 2022. Natural products and drug discovery. National Science Review 9(11), nwac206.
[21] Newman, D.J., Cragg, G.M., 2020. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. Journal of Natural Products 83(3), 770–803.

[22] 21. Nannuri S., Gantla H. R., Ananya D., Vennela A., Bhuvana B., Neha G. (2026). Automated Brain Tumor Segmentationin Mri Via U-Net-Based Deep Neural Networks. International Journal of Engineering Sciences & Research Technology, 15(4), 1–8. https://doi.org/10.64149/j.ijesrt.15.4.1-8

[23] Odugbemi, A.I., Nyirenda, C., Christoffels, A., Egieyeh, S.A., 2024. Artificial intelligence in antidiabetic drug discovery: the advances in QSAR and the prediction of α-glucosidase inhibitors. Computational and Structural Biotechnology Journal 23, 2964–2977.

[24] Ononamadu, C.J., Seidel, V., 2024. Exploring the antidiabetic potential of *Salvia officinalis* using network pharmacology, molecular docking and ADME/drug-likeness predictions. Plants 13(20), 2892.

[25] Ononamadu, C.J., Ben Ahmed, Z., Seidel, V., 2025. Network pharmacology, molecular docking and molecular dynamics studies to predict the molecular targets and mechanisms of action of *Melissa officinalis* phytoconstituents in type-2 diabetes mellitus. Plants 14(18), 2828.

[26] Piñero, J., Ramírez-Anguita, J.M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., Furlong, L.I., 2020. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Research 48(D1), D845–D855.

[27] Sorokina, M., Merseburger, P., Rajan, K., Yirik, M.A., Steinbeck, C., 2021. COCONUT online: Collection of Open Natural Products database. Journal of Cheminformatics 13(1), 2.

[28] Srisongkram, T., Waithong, S., Thitimetharoch, T., Weerapreeyakul, N., 2022. Machine learning and in vitro chemical screening of potential α-amylase and α-glucosidase inhibitors from Thai indigenous plants. Nutrients 14(2), 267.

[29] Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., Bork, P., Jensen, L.J., von Mering, C., 2023. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Research 51(D1), D638–D646.

[30] Szklarczyk, D., Nastou, K., Koutrouli, M., Kirsch, R., Mehryary, F., Hachilif, R., Hu, D., Peluso, M.E., Huang, Q., Fang, T., Doncheva, N.T., Pyysalo, S., Bork, P., Jensen, L.J., von Mering, C., 2025. The STRING database in 2025: protein networks with directionality of regulation. Nucleic Acids Research 53(D1), D730–D737.

[31] Teufel, F., Orgutsova, K., Genitsaridi, I., Carrillo-Larco, R.M., Varghese, J.S., Marcus, M.E., Sacre, J.W., Boyko, E.J., Magliano, D.J., Ali, M.K., 2026. Global, regional, and national estimates of undiagnosed diabetes in adults: findings from the 2025 IDF Diabetes Atlas. Diabetes Care 49(3), 490–496.

[32] Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al., 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research 50(D1), D439–D444.

[33] Vivek-Ananth, R.P., Mohanraj, K., Sahoo, A.K., Samal, A., 2023. IMPPAT 2.0: an enhanced and expanded phytochemical atlas of Indian medicinal plants. ACS Omega 8(9), 8827–8845.

[34] Zdrazil, B., Felix, E., Hunter, F., Manners, E.J., Blackshaw, J., Corbett, S., de Veij, M., Ioannidis, H., Lopez, D.M., Mosquera, J.F., et al., 2024. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Research 52(D1), D1180–D1192.

[35] Zhang, Y., Zhu, Y., Bao, X., Dai, Z., Shen, Q., Wang, L., Xue, Y., 2024. Mining bovine milk proteins for DPP-4 inhibitory peptides using machine learning and virtual proteolysis. Research 7, 0391.