



OPEN ACCESS

EDITED BY

Huixiao Hong,
United States Food and Drug Administration,
United States

REVIEWED BY

Antonina L. Nazarova,
University of Southern California,
United States
Heba Askar,
University of Sadat City, Egypt

*CORRESPONDENCE

Jake Y. Chen,
✉ jakechen@uab.edu

RECEIVED 31 March 2025

ACCEPTED 13 May 2025

PUBLISHED 21 July 2025

CITATION

Hossain D, Saghapour E and Chen JY (2025)
NeSyDPP-4: discovering DPP-4 inhibitors for
diabetes treatment with a neuro-symbolic AI
approach.
Front. Bioinform. 5:1603133.
doi: 10.3389/fbinf.2025.1603133

COPYRIGHT

© 2025 Hossain, Saghapour and Chen. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with
these terms.

NeSyDPP-4: discovering DPP-4 inhibitors for diabetes treatment with a neuro-symbolic AI approach

Delower Hossain^{1,2}, Ehsan Saghapour^{2,3} and Jake Y. Chen^{1,2,3*}

¹Department of Computer Science, The University of Alabama at Birmingham, Birmingham, AL, United States, ²System Pharmacology and AI Research Center (SPARC), The University of Alabama at Birmingham, Birmingham, AL, United States, ³Department of Biomedical Informatics and Data Science, School of Medicine, The University of Alabama at Birmingham, Birmingham, AL, United States

Introduction: Diabetes Mellitus (DM) constitutes a global epidemic and is one of the top ten leading causes of mortality (WHO, 2019), projected to rank seventh by 2030. The US National Diabetes Statistics Report (2021) states that 38.4 million Americans have diabetes. Dipeptidyl Peptidase-4 (DPP-4) is an FDA-approved target for the treatment of type 2 diabetes mellitus (T2DM). However, current DPP-4 inhibitors may cause adverse effects, including gastrointestinal issues, severe joint pain (FDA safety warning), nasopharyngitis, hypersensitivity, and nausea. Moreover, the development of novel drugs and the *in vivo* assessment of DPP-4 inhibition are both costly and often impractical. These challenges highlight the urgent need for efficient *in-silico* approaches to facilitate the discovery and optimization of safer and more effective DPP-4 inhibitors.

Methodology: Quantitative Structure-Activity Relationship (QSAR) modeling is a widely used computational approach for evaluating the properties of chemical substances. In this study, we employed a Neuro-symbolic (NeSy) approach, specifically the Logic Tensor Network (LTN), to develop a DPP-4 QSAR model capable of identifying potential small-molecule inhibitors and predicting bioactivity classification. For comparison, we also implemented baseline models using Deep Neural Networks (DNNs) and Transformers. A total of 6,563 bioactivity records (SMILES-based compounds with IC₅₀ values) were collected from ChEMBL, PubChem, BindingDB, and GTP. Feature sets used for model training included descriptors (CDK Extended-PaDEL), fingerprints (Morgan), chemical language model embeddings (ChemBERTa-2), LLaMa 3.2 embedding features, and physicochemical properties.

Results: Among all tested configurations, the Neuro-symbolic QSAR model (NeSyDPP-4) performed best using a combination of CDK extended and Morgan fingerprints. The model achieved an accuracy of 0.9725, an F1-score of 0.9723, an ROC AUC of 0.9719, and a Matthews correlation coefficient (MCC) of 0.9446. These results outperformed the baseline DNN and Transformer models, as well as existing state-of-the-art (SOTA) methods. To further validate the robustness of the model, we conducted an external evaluation using the Drug Target Common (DTC) dataset, where NeSyDPP-4 also demonstrated strong performance, with an accuracy of 0.9579, an AUC-ROC of 0.9565, a Matthews Correlation Coefficient (MCC) of 0.9171, and an F1-score of 0.9577.

Discussion: These findings suggest that the NeSyDPP-4 model not only delivered high predictive performance but also demonstrated generalizability to external datasets. This approach presents a cost-effective and reliable alternative to traditional *vivo* screening, offering valuable support for the identification and classification of biologically active DPP-4 inhibitors in the treatment of type 2 diabetes mellitus (T2DM).

KEYWORDS

neuro-symbolic artificial intelligence, deep learning, DPP-4, drug discovery, machine learning, QSAR

1 Introduction

Diabetes mellitus (DM) is a chronic metabolic disorder characterized by elevated blood glucose levels, posing a significant global health burden. According to the World Health Organization (WHO) 2019 report, diabetes ranks among the top 10 leading causes of mortality, with an estimated 1.6 million deaths worldwide (World Health Organization: WHO, 2024; World Health Organization, 2020). In the United States, diabetes is a significant public health challenge, affecting approximately 38 million people (11.3% of the population) and leading to \$327 billion in medical expenses and lost wages annually (CDC, 2024). Beyond economic costs, diabetes is associated with severe complications, including blindness, kidney failure, stroke, heart disease, and neuropathy. DM is broadly classified into type 1 diabetes mellitus (T1DM) and type 2 diabetes mellitus (T2DM), with T2DM accounting for over 90% of all cases. One decisive therapeutic target for managing Type 2 Diabetes Mellitus (T2DM) is the Dipeptidyl Peptidase-4 (DPP-4) enzyme, which plays a key role in regulating glucose metabolism. DPP-4 inhibitors, a class of FDA-approved medications, help control blood sugar levels by inhibiting this enzyme. However, current DPP-4 inhibitors have been linked to adverse effects such as gastrointestinal issues, severe joint pain, nasopharyngitis, hypersensitivity, and nausea (Huang et al., 2020). As a result, discovering safer and more effective DPP-4 inhibitors remains a critical research challenge.

Artificial intelligence (AI) has revolutionized diabetes management and drug discovery over the past two decades. Early AI models focused on predicting glucose levels, providing insulin dosage recommendations, and monitoring patients. In recent years, AI has rapidly advanced in the field of *de novo* drug design by leveraging large-scale molecular datasets. These models can not only generate novel drug candidates but also identify repurposable inhibitors and uncover complex relationships among genes, proteins, and disease mechanisms. In the field of DPP-4 inhibitor prediction, quantitative structure–activity relationship (QSAR) models have been widely developed using machine learning techniques such as random forest, support vector machines (SVMs), XGBoost, gradient boosting machines, and deep neural networks (DNNs) (Gong et al., 2021; Hermansyah et al., 2021; Ojo et al., 2021; Septiawan et al., 2022; Bustamam et al., 2021; Ajiboye et al., 2021). Although these models have demonstrated high predictive performance, they have limitations, including poor interpretability, data inefficiency, and a lack of reasoning capabilities. The black-box nature of deep learning models

further complicates their use in critical healthcare applications, where transparency, logical reasoning, and explainability are vital (Hassan et al., 2022).

To address these challenges, neuro-symbolic (NeSy) AI (Hossain and Chen, 2025) has emerged as a promising paradigm that combines neural networks with symbolic reasoning for more interpretable and data-efficient learning. In contrast to traditional AI approaches that rely solely on data, Neuro-symbolic AI (NeSy AI) integrates domain knowledge with data-driven learning, enabling logical reasoning and making it especially well-suited for healthcare and drug discovery applications (Hossain et al., 2025) (Hassan et al., 2022). Studies identified numerous NeSy models that have demonstrated immense success in biomedical applications (Hossain and Chen, 2025; Yu et al., 2023; Wang W. et al., 2024), such as protein function prediction [MultiPredGO (Giri et al., 2020)], gene sequence analysis [KBANN (Towell and Shavlik, 1994)], diabetic retinopathy diagnosis [ExplainDR (Jang et al., 2021)], predicting the structure of proteins [extended KBANN (Maclin and Shavlik, 1994)], cardiotoxicity assessment hERG-LTN (Hossain et al., 2025), (Ontology) RRN (Yang et al., 2017), NSRL (Hohenecker and Lukasiewicz, 2020), Neuro-Fuzzy (Yang et al., 2020), FSKBANN (Kora et al., 2019), DeepMiRGO (Wang et al., 2019), NS-VQA (Yi et al., 2018), DFOL-VQA (Amizadeh et al., 2020), LNN (Riegel et al., 2020), NofM (Towell and Shavlik, 1991), PP-DKL (Lavin, 2022), FSD (Dobosz and Duch, 2008), CORGI (Arabshahi et al., 2021), NeurASP (Shi et al., 2019), XNMs (Teru et al., 2020), Semantic loss (Xu et al., 2018), NS-CL (Mao et al., 2019), and Logic Tensor Networks (LTNs) (Badreddine et al., 2021). In this study, we explore a hybrid neuro-symbolic approach integrating LTNs for DPP-4 bioactivity prediction, aiming to enhance predictive accuracy while enabling logical reasoning for novel drug discovery in T2DM treatment.

The paper's main contributions are summarized as follows:

- 1) We developed a scalable and robust AI predictive model that demonstrates significant improvements in accuracy for predicting the potency of T2DM inhibitors.
- 2) A novel integration of data and rules (Knowledge) for DPP-4 inhibitor bioactivity classification.
- 3) We acquired and utilized a more diverse set of compound datasets, including chemical embeddings, descriptors, fingerprints, and physicochemical properties, which previous studies have not explored. The developed NeSyDPP-4 model can be used to discover novel DPP-4 active drugs by scanning large molecular datasets, such as ZINC, and identifying new candidate compounds, thereby accelerating the *de novo* design of drugs. Additionally,

it facilitates QSAR model downstream applications, including virtual screening, contraindications, bioactivity indications, and other key elements of DPP-4 inhibitor therapy in clinical settings. These applications encompass docking, affinity prediction, ADMET analysis, and molecular dynamics (MD) studies for DPP-4 clinical applications.

The remainder of this manuscript is organized as follows: [Section 2](#) provides the background, offering essential insights into the problem domain and its significance. [Section 3](#) briefly reviews related work, highlighting existing approaches and their contributions to DPP-4 bioactivity classification. [Section 4](#) describes the methodology, detailing the proposed approach, datasets, and algorithms used. [Section 5](#) reports the results obtained from the experimental evaluation. [Section 6](#) offers a discussion, interpreting and comparing the findings with existing studies. Finally, [Section 7](#) concludes the paper with key findings and guidance for future research.

2 Background

2.1 Dipeptidyl peptidase-4 inhibitors

DPP-4 is an enzyme primarily involved in glucose metabolism, particularly regulating blood glucose levels in T2DM. DPP-4 is an FDA-approved target for T2DM treatment. The primary aim of the DPP-4 inhibitor is to prevent the degradation of incretin hormones and improve blood glucose control. Several FDA-approved DPP-4 inhibitors include sitagliptin, saxagliptin, linagliptin, and alogliptin (FDA, 2023; [Supplementary Appendix A](#)). In addition, [Figure 1](#) depicts the 2D structures of FDA-approved DPP-4 inhibitors collected from ChEMBL.

2.2 Quantitative structure–activity relationship

QSAR modeling ([Perkins et al., 2003](#)) is a computational technique that uncovers patterns between molecular features and experimental outcomes, which helps predict the biological activity of compounds based on their chemical composition. Machine learning algorithms like artificial neural networks (ANNs), SVMs, and random forests are commonly applied to build accurate QSAR models. This approach accelerates drug discovery by enabling *in silico* assessment, reducing the reliance on extensive laboratory testing (wet-lab).

2.3 Symbolic AI

Good Old-Fashioned Artificial Intelligence (GOF AI), commonly referred to as Symbolic AI, is a classical AI approach emphasizing knowledge representation and reasoning. It was the dominant paradigm from the 1950s to the 1980s ([Hossain and Chen, 2025](#)), using methods like logic, rules, ontologies, decision trees, and knowledge graphs. Although symbolic AI excels in explanation, interpretability, and structured decision-making, it struggles to perform effectively at scale and with noisy data.

2.4 Sub-symbolic AI

Sub-symbolic AI, often called a “black box” approach, relies on large-scale data and statistical learning rather than explicit rules. Sub-symbolic (connectionist) AI has driven significant advancements in the modern AI era, such as autonomous driving systems and healthcare. For instance, a recent breakthrough contribution is building AlphaFold, a highly accurate protein structure prediction model developed by google DeepMind research. However, ANNs are the heart of this connectionist system. Although it excels in pattern recognition and handling unstructured, noisy, and big data, it lacks transparency, explainability, and reasoning.

2.5 Neuro-symbolic AI

Neuro-symbolic AI ([Hossain and Chen, 2025](#)) is an emerging branch of artificial intelligence constructed to bridge the gap between symbolic and connectionist approaches by integrating their strengths and eliminating their flaws. This hybrid integration aims to create AI systems that are both interpretable and capable of reasoning. Interpretable and reasoning-based AI enhances trust, transparency, and decision-making in healthcare ([Giri et al., 2020](#); [Towell and Shavlik, 1994](#); [Jang et al., 2021](#); [Hossain et al., 2025](#); [Hassan et al., 2022](#)), enabling accurate diagnoses and personalized treatments. It also improves safety, accountability, and regulatory compliance.

2.6 Logic Tensor Network

The LTN is a neuro-symbolic framework developed at Sony Computer Science Laboratories (Sony CSL) that enables querying, learning, and reasoning with complex data and abstract knowledge. It uses a differentiable first-order language called Real Logic to integrate logical reasoning with data-driven learning ([Badreddine et al., 2021](#)). The core advantage of this paradigm is its ability to perform reasoning, which is expressed through logical components, and is highly scalable. Additionally, it offers a comprehensive framework capable of handling supervised and unsupervised tasks, including regression, classification, and clustering. In this study, we conceived this model (LTN) for DPP-4 bioactivity classification, and more details are discussed in the *Methodology* section.

3 Related work

DPP-4 inhibitors are a significant class of compounds used in treating type 2 diabetes; predicting their bioactivity is essential in early drug discovery. Several machine learning-based studies have been explored for classifying DPP-4 inhibitors. For instance, QSAR models have been widely applied with descriptors such as molecular fingerprints and physicochemical properties to predict inhibitory activity. Techniques like random forests, support vector machines, and deep neural networks have shown promising results

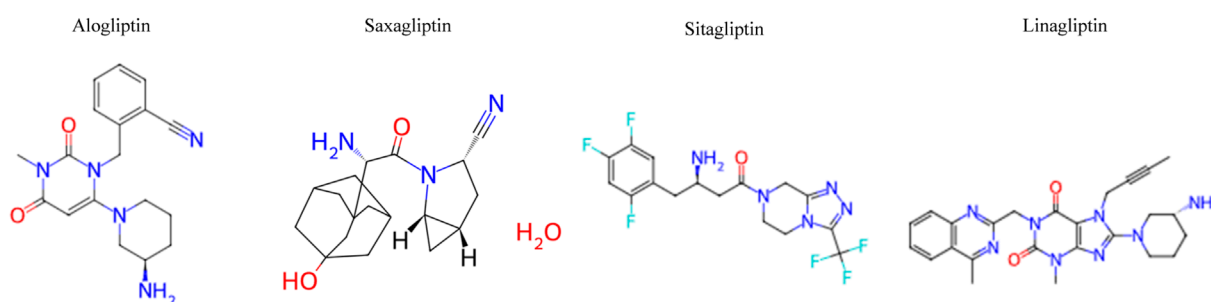


FIGURE 1
2D structure of FDA-approved dipeptidyl peptidase-4 (DPP-4) inhibitors collected from ChEMBL.

in improving classification accuracy and guiding virtual screening processes. For instance, Hermansyah et al. (2020) employed Random Forest and DNNs, achieving an accuracy of 0.9221 using CDK fingerprint and molecular properties. Bustamam et al. (2021) developed a QSAR–DNN model, yielding an accuracy of 0.904. In addition, Cai et al. (2017) applied a Naïve Bayesian (NB) approach using various fingerprint extractions, such as ECFP_4, ECFP_6, FCFP_4, and FCFP_6, reaching an accuracy of 0.872. Ulfa et al. (2021) combined Conv1D and LSTM layers for bioactivity classification, using CatBoost-selected fingerprint features, and achieved an accuracy of 0.8618. Finally, Hermansyah et al. (2023) used XGBoost with CDK and ECFP-6 features, reporting an accuracy of 0.8164. These studies highlight the progression from traditional machine learning models to advanced deep learning approaches, which have steadily improved predictive performance in DPP-4 bioactivity classification. However, to date, neuro-symbolic approaches have not been effectively integrated into DPP-4 research, even though such integration is essential for the discovery of novel chemical compounds. Moreover, previous research lacked experimentation with diverse chemical features, such as chemical language model embeddings and physicochemical properties. To address these gaps, we propose the NeSyDPP-4 strategy, which leverages Logic Tensor Networks and incorporates a wide range of feature representations, including LLaMA3.2 embeddings, PaDELPy CDKExtended fingerprints, Morgan fingerprints, chemical language model embeddings, and physicochemical properties.

4 Methodology

This section describes a set of procedures to determine the performance of LTNs (Badreddine et al., 2021), DNNs, and an advanced language model known as Transformer, using the ChEMBL BindingDB, PubChem, and GTP datasets related to DPP-4 inhibitors. This section covers the entire pipeline, including the material procurement, data preprocessing, feature extraction, simulation environment, network architecture, LTN knowledge-based setting, the training and inferencing phases, and the evaluation metrics used to measure the trained model's performance.

4.1 Data acquisition

The study constructed a new DPP-4 cohort using four publicly available chemical compound databases, namely, ChEMBL (Gaulton et al., 2011), BindingDB (Gilson et al., 2015), PubChem, and GTP. The ChEMBL database contains more than 2 million compounds. We retrieved canonical SMILES related to the DPP-4 inhibitor with the target organism *Homo sapiens* using ID: ChEMBL284 and standard type IC₅₀. The data were extracted using the ChEMBL Python API (chembl_webresource_client). The BindingDB manually uses DPP-4 string keywords (dipeptidyl peptidase-4) from their official site. Subsequently, data from PubChem in CSV format were retrieved using the following link, and GTP data were accessed via the corresponding link. In addition, to assess the model's robustness and generalizability, we collected additional DPP-4-related data from Drug Target Common (DTC) via the provided link for external validation. Curated data can be found in the Data Availability section.

4.2 Data preprocessing and feature extraction

The initial bioactivity datasets comprised various irrelevant features. The curated subsets focused explicitly on the IC₅₀ biological activity values, inhibitor identifiers (such as ChEMBL_ID and BindingDB_ID), and canonical SMILES representations and constrained the target organism to *Homo sapiens*. Notably, numerical IC₅₀ values were reported in nanomolar (nM) units for ChEMBL, BindingDB, and GTP, while PubChem provided values in micromolar (μM) units. All IC₅₀ measurements were harmonized to nanomolar units to standardize the data using the conversion formula $nM = \mu M \times 1000$. Subsequently, pIC₅₀ values were computed from the standardized IC₅₀ values through a logarithmic transformation using log₁₀ (Equation 1) to normalize the distribution. Based on these pIC₅₀ values, compounds were labeled as active or inactive according to established thresholds from prior DPP-4 chemical research (Ulfa et al., 2021). After merging all curated datasets, duplicates were removed based on the canonical SMILES column, and entries with missing values in the ID, SMILES, or IC₅₀ fields were discarded. The final dataset was then split using

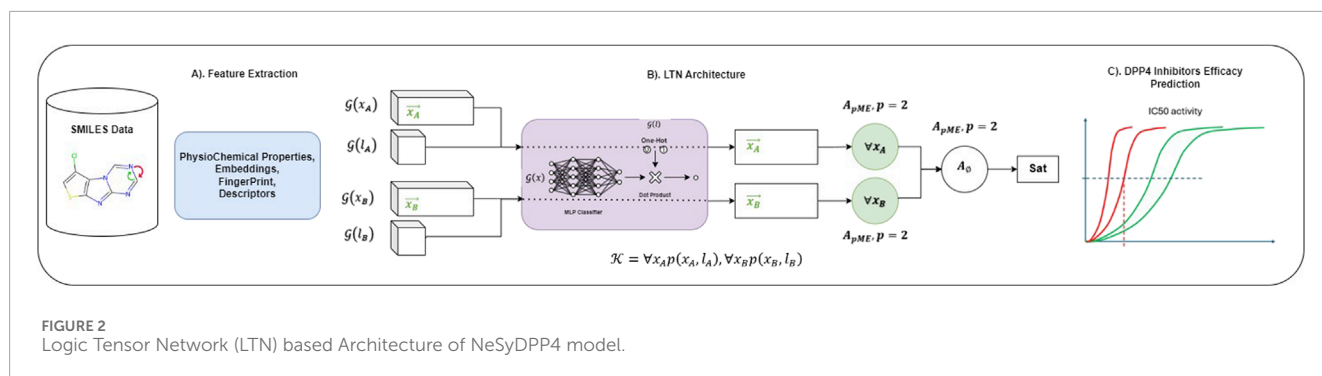


TABLE 1 LTN knowledge-based setting for DPP-4 classification.

Content	Block
Definition of axioms	<ul style="list-style-type: none"> $\forall x_A, p(x_A, l_A)$: all the examples of class A ($Active = 0$) should have a label l_A $\forall x_B, p(x_B, l_B)$: all the examples of class B ($Inactive = 1$) should have a label l_B
Axioms (rules and knowledge base)	$K = \{x_A, p(x_A, l_A), x_B, p(x_B, l_B)\}$
SatAgg is given by	$SatAgg_{\phi \in K} \mathcal{G}_{\theta, x \leftarrow D}(\phi)$
Learning and loss	$L = \left(1 - SatAgg_{\phi \in K} \mathcal{G}_{\theta, x \leftarrow D}(\phi) \right)$

Note: This table was constructed inspired by the official LTN model.

More detail [Supplementary Appendix C](#).

TABLE 2 List of hyperparameters used for NeSyDPP4 (LTN) and DNN models for DPP-4 inhibitor classification.

Parameter	LTN	DNN
Units (input sample/hidden units)	(768,384,192,2)	(768,384,192,2)
Activation	ReLU	ReLU
No. of dense layers	3	3
Dropout	0.25	0.25
Seed	42	42
Batch size	128	128
Training epochs	100	100
Learning rate	0.00001	0.00001
Loss function	LTN pMeanError	Sparse Categorical Crossentropy
Optimizer	Adam	Adam

Note: LTN and DNN models were configured identically in terms of MLP architecture, activation function, and training setup to ensure a fair comparison.

a stratified sampling strategy with the *scikit-learn* package, followed by feature scaling using a standard scaler.

Afterward, a diverse array of features was extracted from SMILES representations, encompassing Morgan fingerprints (512,

1024, and 2048 bits), CDKExtended descriptors using PaDELPy (Yap, 2011), chemical embeddings generated via ChemBERTa-2 (Ahmad et al., 2022) and LLaMA3.2 (Ettaleb et al., 2024) from the Hugging Face model, and a comprehensive set of physicochemical properties [molecular weight, hydrophobicity-LogP, topological polar surface area (TPSA), hydrogen bond donors, hydrogen bond acceptors, and rotatable bonds] extracted using RDKit (Installation, 2024).

$$pIC_{50} = -\log_{10}(IC_{50} \times 10^{-9}). \quad (1)$$

4.3 LTN classification model

We used LTNs (Badreddine et al., 2021) to build the NeSyDPP4 classifier. LTNs combine neural networks with first-order logic, which allows us to perform reasoning over structured knowledge while learning from data. The architecture has two key components: logic and neural networks. The visual architecture of the classification model was adopted from the LTN (shown in Figure 2). The logical mechanism contains a set of axioms or rules (explained in detail in the knowledge-based setting), and reasoning is revealed through those rules/axioms. In our context, Table 1 represents the axioms and the relevant knowledge-based component. Notably, other network configuration parameters can be found in Table 2.

The pMeanError aggregator, as shown in Equation 2,

$$pME(u_1, \dots, u_n) = 1 - \left(\frac{1}{n} \sum_{i=1}^n (1 - u_i)^p \right)^{\frac{1}{p}} \quad p \geq 1, \quad (2)$$

Here, pMeanError is computed through universal quantification (“for all” or $\forall_{x \in A} / \forall_{x \in B}$ as shown Figure 2), which refers to the generalized mean of the deviations with respect to the truth (more detail [link](#)). Further, SatAgg’s all attributes can be defined as below (Table 1).

- **SatAgg**: This stands for “Satisfaction Aggregator,” an operator that aggregates the truth values of the formulas in K (if there is more than one rule).
- $\phi \in K$: This part indicates that ϕ (phi) belongs to the set K . ϕ is often used to represent a predicate.
- $\mathcal{G}(\theta)$: This is denoted by grounding (\mathcal{G}) with parameters θ . θ represents a set of parameters or weights in a model.
- $x \leftarrow D$: D indicates the dataset of all examples (domain).
- B is a mini-batch sampled from D .

However, Figure 2 depicts an architecture composed of three segments. Segment A represents several features used to train the model, while Segment B illustrates the LTN-based classification architecture model, which was conceived from the LTN paper. Specifically, feature-label pairs x_A and I_A (for classes A and B, respectively) are used as input into the MLP. The output is evaluated using a universal \forall quantification equation that influences the loss function and updates model weights via backpropagation, the same as for x_B (all the features related to the B class). The logical formulation follows $P(x_A, I_A)$ and $P(x_B, I_B)$, where p represents the model/MLP and K denotes the complete knowledge expressed in real/first-order logic. For more details, refer to the official LTN GitHub repository.

4.4 Model training and validation phase

LTN, DNN, and Transformer models were trained and tested using TensorFlow 2.15.1 with Python 3.10.16 on the UAB server with an NVIDIA A100 80 GB PCIe GPU, and other dependency packages can be found in the project's GitHub repository under environment.yml. We partitioned the data into 70:10:20 ratios over 100 epochs in the training phase. To optimize model performance, and conducted hyperparameter tuning via Grid Search (Supplementary Appendix B), with multiple training trials. The best configurations, trial 01 and trial 07, achieved the highest accuracy of 0.9726, using ReLU activation and a learning rate of 0.0001, with three layers. In addition to experimenting with the LTN, we conducted the simulation with DNNs and Transformer with Keras integrated to compare LTN performance fairly. Table 2 depicts the network configuration parameters and the simulation notebook (project GitHub), which describes the details.

Furthermore, we conducted external validation by extracting additional 1,045 data samples from the DTC dataset using the best-performing model weights. Notably, there was no duplication or overlap between the external validation data and the training dataset. The following metrics, such as accuracy, F-score (F), ROC AUC score, and Matthews correlation coefficient (MCC), were used to assess the trained model's performance evaluation. Additionally, the confusion matrix (CM) provides a visual representation of misclassified classes (Figure 3).

Equation 3 represents the accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Equation 4 represents the F1-score:

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Equation 5 represents the ROC AUC score:

$$\text{ROCAUC} = \int_0^1 \text{TPR} d(\text{FPR}) \quad (5)$$

Equation 6 represents the MCC:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (6)$$

5 Result

This section outlines the performance of the proposed NeSyDPP-4 model in identifying potential DPP-4 inhibitors by integrating logical rules into a neural network via the LTN architecture. To achieve this, we extracted a diverse set of molecular features from each SMILES/drug representation. These include the Morgan fingerprint, CDK extended descriptors, and embeddings generated from chemical language models such as ChemBERTa-2 and LLaMA3.2 (via the Hugging Face platform). Additionally, physicochemical properties were computed using RDKit to enrich the feature space with interpretable molecular characteristics. This section presents four result grids: Table 3 shows all the features separated and combined as input results for an ablation study; Table 4 exposes the fair comparison with baseline DNN and transformer architecture performance; Table 5 summarizes the model's performance on the external evaluation; and finally, Table 6 presents the benchmarking evaluation. To illustrate, Table 3 depicts the different input performances of NeSyDPP (LTN). The best-performing feature set is combining CDKExtended + ECFP, which yielded the highest accuracy (97.25%), F1-score (97.23%), AUC ROC (97.19%), and MCC (94.46%), while physicochemical features

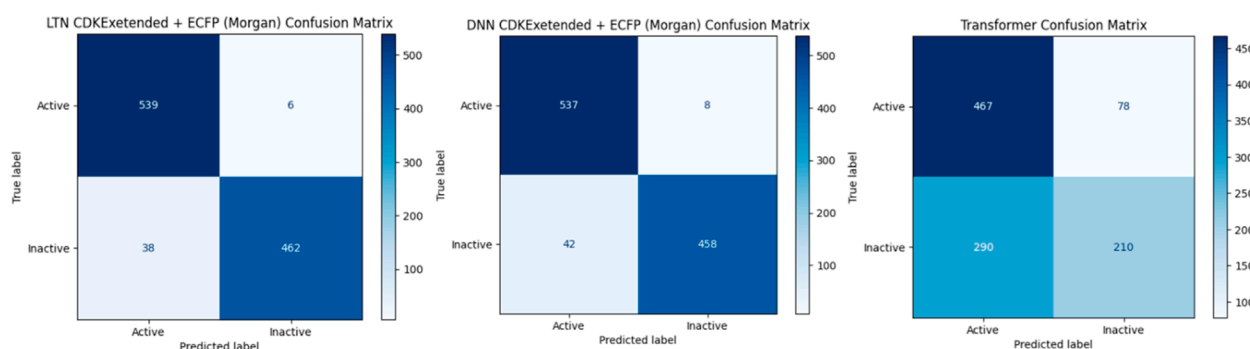


FIGURE 3
Confusion matrix of DNN, Transformer, and LTN using CDKExtended + Morgan features based on DTC dataset evaluation.

TABLE 3 NeSyDPP-4 (LTN) model’s performance comparison using various feature representations and input dimensions for DPP-4 inhibitor classification.

Model	Feature	Input	Acc	F1	AUC ROC	MCC
NeSyDPP-4 (LTN)	CDKExtended + ECFP	1024+(512 + 1024+2048)	0.9725	0.9723	0.9719	0.9446
	ECFP	1024	0.9687	0.9684	0.9680	0.9370
	ECFP	2048	0.9657	0.9654	0.9650	0.9308
	ECFP	512	0.9649	0.9646	0.9643	0.9293
	Combined All	7430	0.9634	0.9631	0.9632	0.9262
	CDKExtended	1024	0.9504	0.9499	0.9492	0.9001
	ChemBERTa-2	768	0.8956	0.8944	0.8935	0.7892
	LLaMA3.2	2048	0.8933	0.8926	0.8933	0.7854
	Physiochemical	6	0.7349	0.7316	0.7309	0.4638

Note: The table reports Accuracy (Acc), F1-score (F1), AUC-ROC, and Matthews Correlation Coefficient (MCC). The model achieved the highest performance (bold) using the combined CDKExtended and ECFP descriptors.

TABLE 4 Comparison of the NeSyDPP-4 (LTN) model with baseline deep learning architecture and the Transformer.

Model	Feature	Input dimension	Acc	F1	AUC ROC	MCC
NeSyDPP-4 (LTN)	CDKExtended + ECFP	1024+(512 + 1024+2048)	0.9725	0.9723	0.9719	0.9446
DNN	CDKExtended + ECFP	1024+(512 + 1024+2048)	0.9695	0.9692	0.9691	0.9385
Transformer	SMILES/Emb	212	0.7821	0.7306	0.8549	0.5641

Note: Models were trained and evaluated on identical feature sets, and results are based on internal data.

TABLE 5 External evaluation (DTC dataset) comparing NeSyDPP-4 with baseline DNN and Transformer model performances.

Model	Feature	Input dimension	Acc	F1	AUC ROC	MCC
NeSyDPP-4 (LTN)	CDKExtended + ECFP	1024+(512 + 1024+2048)	0.9578	0.9576	0.9564	0.9171
DNN	CDKExtended + ECFP	1024+(512 + 1024+2048)	0.9521	0.9518	0.9506	0.9057
Transformer	SMILES/Emb	212	0.6478	0.6251	0.6384	0.3095

TABLE 6 Comparative performance grid shows typical machine learning and deep learning models with NeSyDPP-4 for DPP-4 inhibitor classification results.

Model	Author	Metrics	Result	Reference
NeSyDPP-4	Hossain <i>et al</i>	Accuracy	0.9578	
Random Forest	Oky Hermansyah <i>et al</i>	Accuracy	0.9221	Hermansyah et al. (2020)
DNN	Haris Hamzah <i>et al</i>	Accuracy	0.9060	Hamzah et al. (2020)
QSAR-DNN	Alhadi Bustamam <i>et al</i>	Accuracy	0.9040	Bustamam et al. (2021)
NB	Jie Cai <i>et al</i>	Accuracy	0.8720	Cai et al. (2017)
Conv1D–LSTM	Adawiyah Ulfa <i>et al</i>	Accuracy	0.8618	Ulfa et al. (2021)
XGBoost	Oky Hermansyah <i>et al</i>	Accuracy	0.8164	Hermansyah et al. (2023)

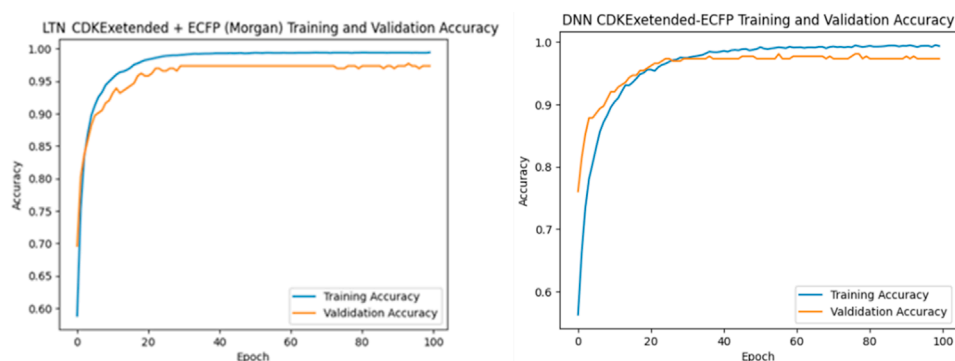


FIGURE 4
The graphs represent epoch and accuracy during the training and validation phases.

alone yield the lowest performance of accuracy (73.49%), F1-score (73.16%), AUC ROC (73.09%), and MCC (46.38%). ChemBERTa-2 and LLaMA3.2 performed comparably but achieved lower performance than the fingerprint-based methods. Overall, physicochemical properties alone are insufficient for effective bioactivity classification.

In addition, Table 4 exhibits the internal validation results comparing the performance of NeSyDPP-4 (LTN), DNN, and Transformer models. The NeSyDPP-4 model, using CDKExtended and ECFP features, achieved the highest performance with 97.25% accuracy and an MCC of 94.46%, highlighting the strength of neuro-symbolic reasoning. The DNN model, using the same input features but without reasoning capabilities, yielded slightly lower results (96.95% accuracy and 93.85% MCC). In contrast, the Transformer model, which relied on SMILES embeddings, exhibited the weakest performance (78.21% accuracy and 56.41% MCC). These findings accentuate the effectiveness of fingerprint-based features over SMILES-based language models for bioactivity classification tasks. Furthermore, in the external evaluation, using the DTC dataset (Table 5), NeSyDPP-4 continued to outperform baseline models, confirming its strong generalization capability.

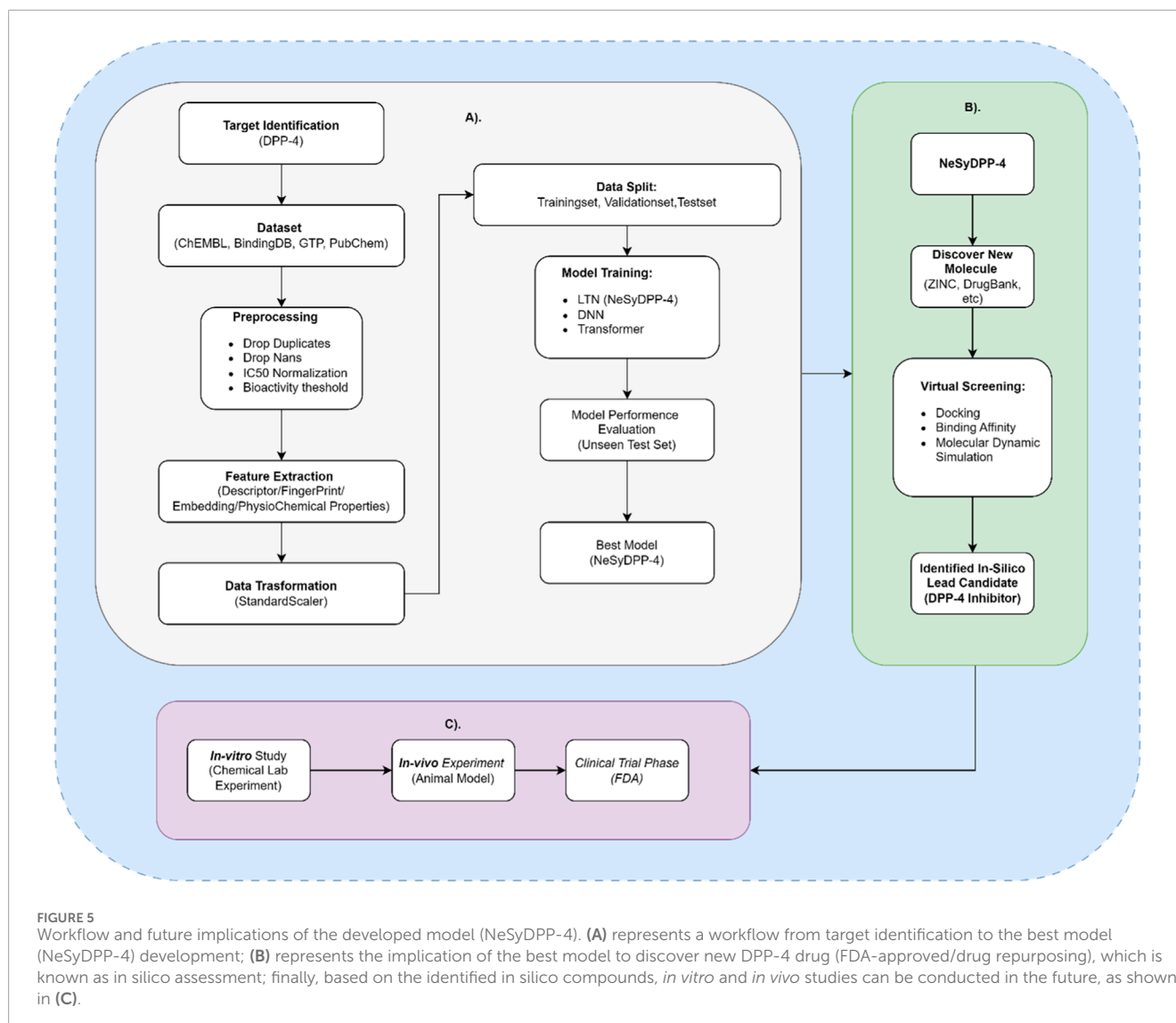
Finally, Table 6 presents a state-of-the-art (SOTA) performance comparison. The proposed NeSyDPP-4 model outperforms all baseline models across accuracy metrics. Traditional models such as Random Forest and XGBoost achieved 92.21% and 81.64% accuracy, respectively, while deep learning models like DNN and QSAR-DNN reached approximately 90%, as reported in prior benchmarks. These results underscore the superior performance of the neuro-symbolic NeSyDPP-4 model. In addition, Figure 4 illustrates the training and validation accuracy curves over 100 epochs for both the DNN and NeSyDPP-4 simulations.

6 Discussion

This article aimed to develop a neuro-symbolic model leveraging LTN, an integration of data and a logic-driven approach, for predicting DPP-4 inhibition in diabetes mellitus. One of the

key challenges in DPP-4 inhibitor research and AI integration is the absence of unified, data and knowledge-driven experimental frameworks that support logical reasoning. Moreover, current approaches fail to effectively utilize diverse features and integrate logical rules, such as extracting embeddings from large language models (LLMs) and physicochemical properties, to predict DPP-4 bioactivity accurately. Furthermore, prior studies have focused on data-driven approaches such as DNN and Transformer. In this context, our study introduces LTN-based NeSyDPP-4, a neuro-symbolic classifier trained on the curated diverse bioactivity activity data and logical rules (Table 1), which demonstrates superior performance in predicting DPP-4 inhibitory activity compared to existing models. Notably, some studies suggest that it can be semi-interpretable since rules (Table 1) are apparently understandable by humans regarding how models should make decisions. However, the study's findings provide valuable insights into the applicability and robustness of the LTN model, discovering bioactivity behavior. To illustrate, the utilization of this advanced machine learning technique (LTN) surpassed the state-of-the-art performance compared to other models with classification tasks; the proposed model demonstrates superior accuracy of 0.9725 and an MCC score of 0.9446 in the internal dataset for DPP-4 inhibitor bioactivity prediction. In contrast, several other studies have reported comparable results: the QSAR-DNN model by Bustamam et al. (2021) achieved an accuracy of 0.9040; Ulfa et al. (2021) reported an accuracy of 0.8618 using Conv1D-LSTM; random forest by Hermansyah et al. (2020) yielded an accuracy of 0.9221; and DNN by Hamzah et al. (2020) obtained an accuracy of 0.9060. Furthermore, the NB model by Cai et al. (2017) gained an accuracy of 0.8720, while ML-based XGBoost by Hermansyah et al. (2023) reported an accuracy of 0.8164.

Overall, this study emphasizes the value of integrating neuro-symbolic modeling, which combines data-driven learning with logical rule-based reasoning, for highly accurate classification of DPP-4 inhibitor activity in the context of diabetes mellitus, a task that conventional data-driven AI approaches (DNN, Transformer) are incapable of. Significantly, the developed NeSyDPP4 model holds substantial promise for future applications, including the discovery of novel DPP-4 compounds, prediction of DPP-4-related



pharmacological interactions, efficient high-throughput screening of molecular libraries to identify potential active agents, and conducting virtual screening based on identified approved/existing substance compounds that can be utilized as drug re-purposing (Figures 5B, C).

6.1 Limitation

Acknowledging the limitations of our study, we state that although the LTN has demonstrated significant promise, it may be incapable of incorporating external, diverse, comprehensive biological additional knowledge with neural networks due to structural limitations.

7 Conclusion

Diabetes mellitus is a vital global health concern, and discovering effective chemical compounds is decisive to tackling this epidemic. In this study, we develop a QSAR system to identify a therapeutic potential compound of DPP-4 inhibitors using an advanced AI framework called LTN that integrates domain-specific knowledge into neural networks. The study is a pioneer in applying the neuro-symbolic strategy in the DM domain and provides new insights that reveal a higher performance for DPP-4 bioactivity classification. The root cause of achieving such performance could be upholding learning and reasoning principles and training neural networks with rules. Furthermore, we experimented with DNN, an NLP Transformer model, whereas

the LTN-based model NeSyDPP-4-QSAR attained the highest accuracy of those baselines' approaches and prior SOTA strategies. In conclusion, the findings of this study prove that the neuro-symbolic approach for uncovering potential DPP-4 inhibitors is promising. However, an ideal direction for future work could involve integrating additional potential neuro-symbolic strategies, such as Semantic Loss and DeepProbLog, to study GLP-1, IDO, and PTP1B inhibitors, which would include extracting a variety of new descriptors and fingerprints from different datasets (PubChem and Drug Bank), focusing on regression tasks.

Data availability statement

The dataset utilized in this study can be found at: <https://drive.google.com/file/d/1SGiYOyuSiirueZR3F6K0d7aHcPT43PYw/view>, and the experimental code repository can be found at: <https://github.com/hossain013/NeSyDPP4-QSAR>.

Author contributions

DH: conceptualization, data curation, formal analysis, investigation, methodology, resources, software, supervision, validation, visualization, writing – original draft, and writing – review & editing. ES: conceptualization, formal analysis, investigation, resources, validation, and writing – review & editing. JC: funding acquisition, project administration, resources, supervision, and writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The work is partly supported by an NIH grant R21DK129968 and research startup funding awarded to Jake Chen.

References

- Ahmad, W., Simon, E., Chithrananda, S., Grand, G., and Ramsundar, B. (2022). ChemBERTa-2: towards chemical foundation models. *arXiv*. doi:10.48550/arXiv.2209.01712
- Ajiboye, B. O., Iwaloye, O., Owolabi, O. V., Ejeje, J. N., Okerewa, A., Johnson, O. O., et al. (2021). Screening of potential antidiabetic phytochemicals from *Gongronema latifolium* leaf against therapeutic targets of type 2 diabetes mellitus: multi-targets drug design. *SN Appl. Sci.* 4 (1), 14. doi:10.1007/s42452-021-04880-2
- Amizadeh, S., Palangi, H., Polozov, O., Huang, Y., and Koishida, K. (2020). Neuro-symbolic visual reasoning: disentangling “visual” from “reasoning”. *Int. Conf. Mach. Learn.* 1, 279–290. doi:10.48550/arxiv.2006.11524
- Arabshahi, F., Lee, J., Gawarecki, M., Mazaitis, K., Azaria, A., and Mitchell, T. (2021). Conversational Neuro-Symbolic commonsense reasoning. *Proc. AAAI Conf. Artif. Intell.* 35 (6), 4902–4911. doi:10.1609/aaai.v35i6.16623
- Badreddine, S., Garcez, A. D., Serafini, L., and Spranger, M. (2021). Logic tensor networks. *Artif. Intell.* 303, 103649. doi:10.1016/j.artint.2021.103649
- Bustamam, A., Hamzah, H., Husna, N. A., Syarifina, S., Dwimantara, N., Yanuar, A., et al. (2021). Artificial intelligence paradigm for ligand-based virtual screening on the drug discovery of type 2 diabetes mellitus. *J. Big Data* 8 (1), 74. doi:10.1186/s40537-021-00465-3
- Cai, J., Li, C., Liu, Z., Du, J., Ye, J., Gu, Q., et al. (2017). Predicting DPP-IV inhibitors with machine learning approaches. *J. Computer-Aided Mol. Des.* 31 (4), 393–402. doi:10.1007/s10822-017-0009-6
- CDC (2024). Methods for the national diabetes statistics report. *Diabetes*. Available online at: https://www.cdc.gov/diabetes/php/data-research/methods.html?CDC_AAref_Val=https://www.cdc.gov/diabetes/data/statistics-report/index.html
- Dobosz, K., and Duch, W. (2008). “Fuzzy symbolic dynamics for neurodynamical systems,” in *Lecture notes in computer science*, 471–478. doi:10.1007/978-3-540-87559-8_49
- Ettaleb, M., Kamel, M., Moriceau, V., and Aussenac-Gilles, N. (2024). The Llama 3 Herd of Models. *Arxiv*. doi:10.48550/arxiv.2407.21783
- FDA (2023). FDA approved dipeptidyl peptidase IV (DPP IV) inhibitors. Available online at: <https://www.ncbi.nlm.nih.gov/books/NBK542331/#:~:text=DPP%2D4%20inhibitors%2C%20known%20as,saxagliptin%2C%20linagliptin%2C%20and%20alogliptin>
- Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., et al. (2011). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40 (D1), D1100–D1107. doi:10.1093/nar/gkr777
- Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., and Chong, J. (2015). BindingDB in 2015: a public database for medicinal chemistry, computational

Acknowledgments

The authors acknowledge the biomedical data science infrastructure and staff support provided by the UAB U-BRITE program.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2025.1603133/full#supplementary-material>

chemistry and systems pharmacology. *Nucleic Acids Res.* 44 (D1), D1045–D1053. doi:10.1093/nar/gkv1072

Giri, S. J., Dutta, P., Halani, P., and Saha, S. (2020). MulTiPREDGO: deep Multi-Modal Protein function prediction by amalgamating protein structure, sequence, and interaction information. *IEEE J. Biomed. Health Inf.* 25 (5), 1832–1838. doi:10.1109/jbhi.2020.3022806

Gong, J. N., Zhao, L., Chen, G., Chen, X., Chen, Z. D., and Chen, C. Y. (2021). A novel artificial intelligence protocol to investigate potential leads for diabetes mellitus. *Mol Divers.* 25, 1375–1393. doi:10.1007/s11030-021-10204-8

Hamzah, H., Bustamam, A., Yanuar, A., and Sarwinda, D. (2020). Predicting the molecular structure relationship and the biological activity of DPP-4 inhibitor using deep neural network with CatBoost method as feature selection. *IEEEExplore*, 101–108. doi:10.1109/icacsis51025.2020.9263204

Hassan, M., Guan, H., Melliou, A., Wang, Y., Sun, Q., Zeng, S., et al. (2022). Neuro-symbolic learning: principles and applications in ophthalmology. arXiv preprint arXiv:2208.00374.

Hermansyah, O., Bustamam, A., and Yanuar, A. (2020). Virtual screening of DPP-4 inhibitors using QSAR-Based artificial intelligence and molecular docking of HIT compounds to DPP-8 and DPP-9 enzymes. *Res. Square Res. Square*. doi:10.21203/rs.2.22282/v2

Hermansyah, O., Bustamam, A., and Yanuar, A. (2021). Virtual screening of dipeptidyl peptidase-4 inhibitors using quantitative structure–activity relationship-based artificial intelligence and molecular docking of hit compounds. *Comput. Biol. Chem.* 95, 107597. doi:10.1016/j.compbiolchem.2021.107597

Hermansyah, O., Rahmawati, S., Dwi Putri Masrijal, C., and Intan Perma Sari, R. (2023). *Identification of DPP-4 inhibitor active compounds using machine learning classification*. Canada: International Journal of Chemical and Biochemical Sciences.

Hohenecker, P., and Lukasiewicz, T. (2020). Ontology reasoning with deep neural networks. *J. Artif. Intell. Res.* 68. doi:10.1613/jair.1.11661

Hossain, D., and Chen, J. Y. (2025). A study on neuro-symbolic artificial intelligence: healthcare perspectives. arXiv preprint arXiv:2503.18213.

Hossain, D., Chen, J. Y., and Abir, F. A. (2025). “hERG-LTN: a new paradigm,” in *hERG cardiotoxicity assessment using neuro-symbolic and generative AI embedding (MegaMolBART, Llama3. 2, gemini, DeepSeek) approach* (bioRxiv USA: Cold Spring Harbor Laboratory), 2025–2102.

Huang, J., Jia, Y., Sun, S., and Meng, L. (2020). Adverse event profiles of dipeptidyl peptidase-4 inhibitors: data mining of the public version of the FDA adverse event reporting system. *BMC Pharmacol. Toxicol.* 21 (1), 68. doi:10.1186/s40360-020-00447-w

Installation (2024). The RDKit 2024.09.6 documentation. Available online at: <https://www.rdkit.org/docs/Install.html>.

Jang, S.-I., Girard, M. J. A., and Thiery, A. H. (2021). *Explainable diabetic Retinopathy classification based on neural-symbolic learning*. arXiv (Cornell University), 104–114. Available online at: <http://ceur-ws.org/Vol-2986/paper8.pdf>.

Kora, P., Meenakshi, K., Swaraja, K., Rajani, A., and Islam, M. K. (2019). Detection of Cardiac arrhythmia using fuzzy logic. *Inf. Med. Unlocked* 17, 100257. doi:10.1016/j.imu.2019.100257

Lavin, A. (2022). “Neuro-Symbolic neurodegenerative disease modeling as probabilistic programmed deep kernels,” in *Studies in computational intelligence*, 49–64. doi:10.1007/978-3-030-93080-6_5

Maclin, R., and Shavlik, J. W. (1994). “Refining algorithms with knowledge-based neural networks: improving the Chou-Fasman algorithm for protein folding,” in *Conference on Learning Theory*, 249–286. Available online at: <http://dl.acm.org/citation.cfm?id=188535>.

Mao, J., Gan, C., Kohli, P., Tenenbaum, J. B., and Wu, J. (2019). The Neuro-Symbolic Concept Learner: interpreting scenes, words, and sentences from natural supervision. *Int. Conf. Learn. Represent.* doi:10.48550/arxiv.1904.12584

Ojo, O. A., Ojo, A. B., Okolie, C., Abdurrahman, J., Barnabas, M., Egbomwan, I. O., et al. (2021). Elucidating the interactions of compounds identified from Aframomum melegueta seeds as promising candidates for the management of diabetes mellitus: a computational approach. *Inf. Med. Unlocked* 26, 100720. doi:10.1016/j.imu.2021.100720

Perkins, R., Fang, H., Tong, W., and Welsh, W. J. (2003). Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. *Environ. Toxicol. Chem.* 22 (8), 1666–1679. doi:10.1897/01-171

Riegel, R., Gray, A. G., Luus, F. P. S., Khan, N., Makondo, N., Akhalwaya, I. Y., et al. (2020). Logical neural networks. *arXiv Cornell Univ.* doi:10.48550/arxiv.2006.13155

Septiawan, N. R. R., Prakoso, N. B. H., and Kurniawan, N. I. (2022). DPP IV inhibitors activities prediction as an anti-diabetic agent using particle swarm optimization-support vector machine method. *J. RESTI Rekayasa Sist. Dan. Teknol. Inf.* 6 (6), 974–980. doi:10.29207/resti.v6i6.4470

Shi, J., Zhang, H., and Li, J. (2019). “Explainable and explicit visual reasoning over scene graphs,” in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 8368–8376. doi:10.1109/cvpr.2019.00857

Teru, K., Denis, E., and Hamilton, W. (2020). Inductive relation prediction by subgraph reasoning. *Int. Conf. Mach. Learn.* 1, 9448–9457. doi:10.48550/arxiv.1911.06962

Towell, G., and Shavlik, J. W. (1991). Interpretation of artificial neural networks: mapping knowledge-based neural networks into rules. *Neural Inf. Process. Syst.* 4, 977–984. doi:10.5555/2986916.2987036

Towell, G. G., and Shavlik, J. W. (1994). Knowledge-based artificial neural networks. *Artif. Intell.* 70 (1–2), 119–165. doi:10.1016/0004-3702(94)90105-8

Ulfa, A., Bustamam, A., Yanuar, A., Amalia, R., and Anki, P. (2021). Model QSAR classification using Conv1D-LSTM of dipeptidyl peptidase-4 inhibitors. *IEEEExplore*, 1–6. doi:10.1109/aims52415.2021.9466083

Wang, D., Jin, J., Li, Z., Wang, Y., Fan, M., Liang, S., et al. (2024). StructuralDPPIV: a novel deep learning model based on atom structure for predicting dipeptidyl peptidase-IV inhibitory peptides. *Bioinformatics* 40 (2), btae057. doi:10.1093/bioinformatics/btae057

Wang, J., Zhang, J., Cai, Y., and Deng, L. (2019). DEEPMIR2GO: inferring functions of human MicroRNAs using a deep Multi-Label Classification model. *Int. J. Mol. Sci.* 20 (23), 6046. doi:10.3390/ijms20236046

Wang, W., Yang, Y., and Wu, F. (2024). Towards data-and knowledge-driven AI: a survey on neuro-symbolic computing. *IEEE Trans. Pattern Analysis Mach. Intell.* 47, 878–899. doi:10.1109/tpami.2024.3483273

Wikipedia (2018). Wikipedia DPP-4 inhibitors. Available online at: https://en.wikipedia.org/wiki/Dipeptidyl_peptidase-4_inhibitor.

World Health Organization (2020). World health statistics 2020: monitoring health for the SDGs, sustainable development goals. Available online at: <https://iris.who.int/bitstream/handle/10665/332070/9789240005105-eng.pdf>.

World Health Organization: WHO (2024). The top 10 causes of death. Available online at: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death#:~:text=Of%20the%2056.9%20million%20deaths%20worldwide%20in%202016%2C,of%20death%20globally%20in%20the%20last%2015%20years>.

Xu, J., Zhang, Z., Friedman, T., Liang, Y., and Broeck, G. (2018). *A semantic loss function for deep learning with symbolic knowledge*. PMLR. Available online at: <https://proceedings.mlr.press/v80/xu18h.html>.

Yang, F., Yang, Z., and Cohen, W. W. (2017). *Differentiable learning of logical rules for knowledge base reasoning*. USA: arXiv (Cornell University). Available online at: https://proceedings.neurips.cc/paper_files/paper/2017/hash/0e55666a4ad822e0e34299df3591d979-Abstract.html.

Yang, Z., Ishay, A., and Lee, J. (2020). “NeurASP: embracing neural networks into answer set programming,” in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*, 1755. Available online at: <https://www.ijcai.org/proceedings/2020/0243.pdf>.

Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* 32 (7), 1466–1474. doi:10.1002/jcc.21707

Yi, K., Wu, J., Gan, C., Torralba, A., Kohli, P., and Tenenbaum, J. B. (2018). Neural-symbolic VQA: disentangling reasoning from vision and language understanding. *arXiv Cornell Univ.* 31, 1031–1042. Available online at: <http://arxiv.org/pdf/1810.02338.pdf>.

Yu, D., Yang, B., Liu, D., Wang, H., and Pan, S. (2023). A survey on neural-symbolic learning systems. *Neural Netw.* 166, 105–126. doi:10.1016/j.neunet.2023.06.028