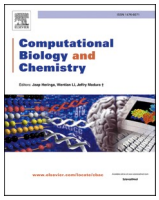
Computational Biology and Chemistry 117 (2025) 108423



Contents lists available at ScienceDirect

Computational Biology and Chemistry

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





Application of Machine Learning (ML) approach in discovery of novel drug targets against *Leishmania*: A computational based approach

Hayat Ali Shah [a,b], Sabina Yasmin [c], Mohammad Yousuf Ansari [d,e,*]

[a] *Wuhan University School of Computer Science Institute of Artificial Intelligence, China*
[b] *National University of Science and Technology, School of Natural Science, Department of Mathematics, Islamabad-44230 Pakistan*
[c] *Department of Pharmaceutical Chemistry, College of Pharmacy, King Khalid University, Abha, Saudi Arabia*
[d] *MM College of Pharmacy, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana 133207, India*
[e] *Ibne Seena College of Pharmacy, Azmi Vidya Nagri, Anjhi Shahabad, Hardoi - Uttar Pradesh (U.P.) 241124 India*

ARTICLE INFO

Keywords:
Leishmania
Molecule
Drugs
Machine learning
Deep learning

ABSTRACT

Molecules with potent anti-leishmanial activity play a crucial role in identifying treatments for leishmaniasis and aiding in the design of novel drugs to combat the disease, ultimately protecting individuals and populations. Various methods have been employed to represent molecular structures and predict effective anti-leishmanial molecules. However, each method faces challenges and limitations that must be addressed to optimize the drug discovery and design process. Recently, machine learning approaches have gained significant importance in overcoming the limitations of traditional methods across various fields. Therefore, there is an urgent need to build a computational pipeline using advanced machine learning and deep learning methods that help to predict anti-leishmanial activity of drug candidates. The proposed pipeline in this paper involves data collection, feature extraction, feature selection and prediction techniques. This review presents a comprehensive computational pipeline for anti-leishmanial drug discovery, highlighting its strengths, limitations, challenges, and future directions to improve treatment for this neglected tropical disease.

1. Introduction

Leishmaniasis is a parasitic disease caused by a parasitic organism of the genus *Leishmania* and transmitted to humans by infected female sandflies. Leishmaniasis is a serious threat to the health of hundreds of thousands of people worldwide, especially in tropical and subtropical regions. The World Health Organization estimates that there are about 700,000–1 million new cases and more than 20,000–30,000 deaths each year. The disease has three clinical manifestations: visceral leishmaniasis, cutaneous leishmaniasis and mucocutaneous leishmaniasis. The nature and outcome of each leishmaniasis varies.

Visceral leishmaniasis is the most dangerous because it has a 95 % mortality rate once it is incurable. Visceral leishmaniasis is a systemic infection characterized by fever, weight loss, splenomegaly and hepatomegaly (Roatt et al., 2014). Even when treated, the mortality rate may be as high as 10 % in endemic areas due to late diagnosis and the development of drug resistance. On the other hand, although cutaneous leishmaniasis is not fatal, it is highly prevalent because of the development of disfiguring skin lesions that lead to social stigma and reduced quality of life (Abadías-Granado et al., 2021). Mucosal leishmaniasis, although rare, can severely damage oral and nasal tissues, leading to long-term disability (Alvar et al., 2012). Scientists are working to develop drugs or treatments that not only cure leishmaniasis, but also reduce serious health consequences (Knight et al., 2023; Okwor and Uzonna, 2016). This will help reduce the burden on patients and communities, especially in areas where leishmaniasis is widespread and poses a major public health challenge.

In this regard, there are various approaches to screen molecules for potent anti-leishmanial activity. One of these approaches consists of in vitro studies, which are the basic method for assessing the direct action of a molecule against *Leishmania* protozoa or other diseases such as Alzheimer's disease under controlled conditions, targeting both the intracellular and extracellular stages of the parasite (Upegui Zapata et al., 2020; Zhang et al., 2024). For example, the ability of miltefosine to kill Leishmania protozoa and inhibit intracellular non-protozoa in in-vitro assays has been extensively studied (Pérez-Victoria et al., 2006). Another example is amphotericin B, which has been evaluated for its activity against *Leishmania braziliensis* and *Leishmania donovani* (Purkait

* Corresponding author.
E-mail address: yousufniper@gmail.com (M.Y. Ansari).

https://doi.org/10.1016/j.compbiolchem.2025.108423
Received 26 October 2024; Received in revised form 6 January 2025; Accepted 7 March 2025
Available online 12 March 2025
1476-9271/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

et al., 2012). More commonly used high-throughput techniques are drug susceptibility testing and cell viability assays, which are useful for identifying potential inhibitors that target specific enzymes or pathways in parasites (Mirza et al., 2011). These studies provide the necessary preliminary data on the cytotoxicity and efficacy of drug candidates. However, *in vitro* experiments have limitations as they occur in isolated systems, are not physiologically relevant, do not depict the full complexity of the immune system, and do not demonstrate the entire life cycle of an organism (Shi et al., 2019). In addition, such studies are resource-intensive and time-consuming, requiring the screening of large libraries of compounds, which may take weeks or months to complete, and therefore require specialists and technical equipment.

In contrast, *In-vivo* studies use animal model's like hamsters or mice to simulate human disease offering insights into the pharmacokinetic, pharmacodynamics and systemic effects of prospective therapies (Ruggeri et al., 2014; Li et al., 2024; Yin et al., 2024; Bulté et al., 2021). For instance, miltefosine, which has been tested in mice models to assess its impact on Leishmania infantum and its ability to modulate immune responses. These studies help assess host-pathogen interactions, immune responses, and overall therapeutic potential (Yardley and Croft, 2000). Another example is liposomal amphotericin B has shown reduced toxicity and improved efficacy compared to conventional amphotericin B formulations when tested in *L. donovani* infected hamsters (Colby et al., 2017). However, In-vivo studies come with several challenges: First In-vivo studies are costly and ethically challenging due to concerns about animal welfare and the repeated infections and treatments required. Second, these studies are complex, long study duration, requiring specialized facilities and careful monitoring, sometimes exhibiting variable outcomes due to differences in animal models (Andersen and Winter, 2019; Francesconi et al., 2023). Third, In-vivo studies provide crucial data on how therapies carry out in a living organism and help bridge the gap between In-vitro real-world therapeutic outcomes and efficacy.

Computational methods such as QSAR (Quantitative Structure-Activity Relationship) modelling and molecular docking allow overcoming in vitro and in vivo data gaps (Cunningham and Fairlamb, 1995). To understand anti-leishmanial activity, QSAR modelling has been performed for many types of leishmanial compounds: Imidazole derivatives have been analyzed based on their lipophilicity as well as functional groups, such as nitroxides and halogens. The results of these analyses have been used to predict the activity of analogues of this class of compounds against *L. major* (Beverley et al., 1986). Molecular docking studies have identified potential inhibitors of *Leishmania* protozoa enzymes, such as tryptic thioredoxin reductase and thymidylate synthase, which are involved in parasite oxidative stress management and DNA(Deoxyribonucleic Acid) synthesis, respectively (Sebastián-Pérez et al., 2022; Turcano et al., 2018). Following this, researchers evaluated the efficacy of platinum or copper-based molecular organometallic complexes against *L. donovani* and linked the ligands structure and metal coordination to the inhibitory effect of the parasite. The substitution pattern of methoxy or hydroxyl groups in naphthoquinone derivatives was also investigated for their activity against *L. infantum* (Majumder et al., 2023). Virtual screening has also been used to identify potential anti-Leishmania compounds by screening large libraries of molecules for their ability to bind to important enzymes in *Leishmania*. For instance, screening can identify inhibitors of *Leishmania* thioredoxin reductase, dihydrofolate reductase thymidylate synthase, or *Leishmania* protease, such as cyclic peptides targeting cathepsin L (Seifert, 2011). This method helps quickly pinpoint promising drug candidates that can disrupt essential parasite functions to accelerates the drug discovery process.

An integrated approach of *in vitro* screening, *in vivo* validation, and the use of computational tools like molecular docking and QSAR is one of the best approaches with which to seek new compounds with anti-Leishmanial activity, and therapeutically, they would have very high potential (Bora et al., 2023; Rabaan et al., 2023; Ferreira et al., 2015; Agu et al., 2023). Hence, this multiple approach will give better speed in the discovery of promising candidates for drugs due to exploiting the synergistic strengths of experimental techniques and *in silico* tools (Ferreira et al., 2015; Wishart et al., 2008). However, the combined approach also has some degree of limitations. Experimental testing on a large compound library is laborious, time-consuming, and costly. Animal testing was ethically problematic and not directly translatable into the human setting (Kim et al., 2016). Computational techniques are only as good as the data used as input. Docking and QSAR are predictions that must be confirmed experimentally. There are difficulties, disadvantages, and challenges with all these techniques, which should not be underestimated.

The aforementioned methods highlight several challenges, particularly in data preparation, which can be both time-consuming and complex. New researchers in this era face difficulties in accessing appropriate models and formatting data for effective prediction and analysis. Therefore, this review article provides a comprehensive overview of data representation techniques, along with their limitations and challenges. It also presents models that facilitate easier access to specific types of data, enabling more efficient and accurate analysis. This review also categorizes various machine learning and deep learning methods, highlighting their limitations and suggesting future research directions to overcome existing challenges and enhance predictive performance.

2. Anti-leishmania molecular prediction computational pipeline

The Leishmaniasis molecular prediction pipeline consisting of several steps: initially, collects molecular data from public databases such as DrugBank (Zdrazil et al., 2024), PubChem (Baranwal et al., 2019), and ChEMBL (Shah et al., 2024), which provide comprehensive information about anti-leishmaniasis molecules, their biological activities, and molecular structures. Subsequently, molecular data are then represented using descriptor methods, graph-based representations, or molecular fingerprints to convert the chemical structure into a numerical format compatible with machine learning models (Yang et al., 2022; Shah et al., 2022; Liu et al., 2024). Next, feature selection is performed to select relevant chemical and pharmacological features to improve model performance and reduce dimensionality (Sifontes-rodríguez et al.). Then, various machine learning methods, such as support vector machines, random forests, gradient boosting, and deep learning models, have been applied to develop predictive models for molecular features related to anti-leishmanial activity. Afterwards, model parameters are optimized by grid search or random search to adjust hyper parameters such as depth, learning rate, or regularization. Finally, model validation is performed using cross-validation and evaluation metrics such as accuracy, precision, recall, f1-score, AUPR and AUROC to ensure that the model generalizes well to unseen data enabling the identification of promising molecules for further experimental testing and optimization. The overall process is shown in Fig. 1, and discussed in detail in the coming subsections, where each step of the pipeline will be thoroughly explained.

2.1. Data curation

Data curation for machine learning begins with data collection, where compound data is extracted from sources like PubChem (Baranwal et al., 2019), such as caffeine, represented by its SMILES string CN1C=NC2 =C1C(=O)N(C(=O)N2C)C and molecular weight. The next step is data cleaning, which involves correcting malformed SMILES strings or filling in missing values. In the data organization phase, the data is structured into tables that list compounds, molecular weights, and activity scores. Feature engineering follows, where relevant descriptors, such as molecular weight and logP, are calculated and selected for predicting enzyme inhibition. During data combination, data from sources like (Zdrazil et al., 2024) or PubChem (Baranwal et al., 2019) is integrated, incorporating additional biological activity

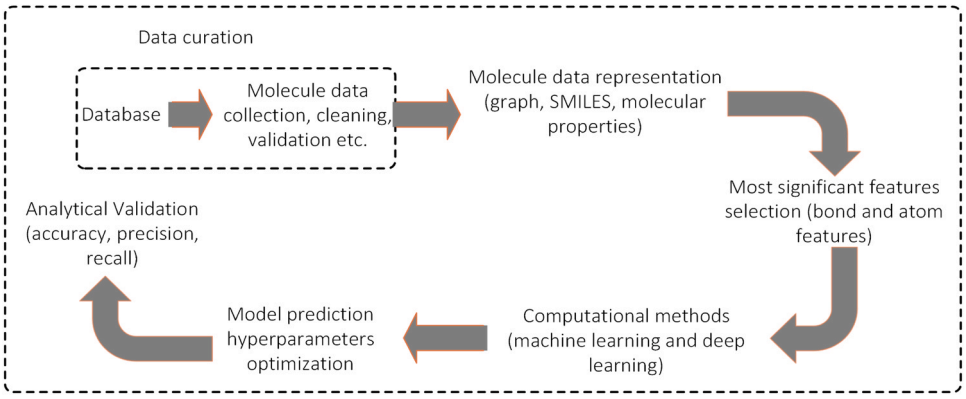


Fig. 1. Workflow for predicting anti-Leishmanial molecules.

information to ensure compatibility with existing datasets. Data transformation prepares the data by coding categorical variables and scaling numerical features, e.g., converting chemical categories to single-click coding vectors. Data validation involves cross-checking a sample of data entries against the original data source to ensure accuracy and consistency. Finally, data storage and access entailed saving the collated dataset to a CSV file and documenting the collation process for future use. Throughout the data collation process, the database played an important role in providing the structured and detailed data necessary for accurate analysis and informed decision-making.

2.1.1. *Molecule databases*

In drug discovery for leishmaniasis, several general data resources play a pivotal role, including PubChem (Baranwal et al., 2019), DrugBank (Zdrazil et al., 2024), BindingDB (Shah et al., 2023) and ChEMBL (Shah et al., 2024), which provide extensive databases of drug-target interactions, bioactivity data and chemical compounds. These platforms are commonly used for QSAR modeling,drug repurposing, drug delivery, and virtual screening allowing researchers to identify potential drug candidates that could be efficient against *Leishmania* (Anuntasomboon et al., 2024). In addition, resources like the BindingDB and Protein Data Bank (PDB) provide binding affinity and structural data, essential for modeling drug target interactions, while Kyoto Encyclopedia of Genes and Genomes (KEGG) offers metabolic pathway, metabolic modules data to understand how drugs interact within the parasites biological systems (Doyle et al., 2009). These databases serve as important starting points for discovering new treatments or repurposing drugs for leishmaniasis.

There are also various specialized data resources related to leishmaniasis research. The Leishmania Genome Project provides genomic data to help study drug resistance and identify new drug targets (Ochoa et al., 2016), while the LeishCyc database (Maeda, 2015) provides metabolic pathway data for *Leishmania* species. The Leishmania drug screening dataset (DSD) brings together molecular bioactivity data tested against *Leishmania*, thus providing a collated resource for drug efficacy studies (Breslin and Pham, 2023). The data contained herein are complementary to other broader drug discovery resources and provide the basis for complex computational methods to identify effective treatments for leishmaniasis.

2.1.1.1. Challenges in accessing and extracting data. This includes incomplete or incorrect information, such as missing SMILES strings, incorrect bioactivity annotations, etc. Other issues include data overlap in most databases; this makes unique screening of compounds cumbersome (O'Boyle et al., 2011). Another challenge is inconsistent use of data formats, as different databases have their own formats and naming conventions. This complicates the process of integrating information from different sources. In addition, unambiguously identifying molecules associated with leishmaniasis can also be challenging, as many compounds are annotated as having broad anti-parasitic activity but are not associated with a specific disease.

2.1.1.2. Solutions to overcome challenges. These will require a good data cleaning pipeline to correct incorrectly formatted SMILES strings, remove duplicate data, and cross-validate molecular properties against multiple sources to further improve quality. This may also involve the use of RDKit (Gaulton et al., 2012) and Open Babel (Hu et al., 2022), which help to convert molecular characterization into a consistent format compatible with different platforms. These machine learning algorithms classify and annotate compounds based on biological activity and disease relevance, thus reducing manual sorting efforts. Finally, a disease-specific database has been developed to improve the relevance and reliability of leishmaniasis research by integrating collated data from general repositories such as ChEMBL and literature sources (Zhu et al., 2023; Shah et al., 2024).

2.2. *Molecular representation*

Molecular representations are methods of representing the properties and structures of molecules. These formats are necessary in the fields of biochemistry, chemistry, pharmacology and related fields as they help in predicting chemical reactions, understanding the behavior of molecules, designing drugs, etc. In the case of leishmaniasis, Covid-19 and leukemia or kind of diseases, molecular representation refers to the display of molecules in various formats that may be required for drug development or related research (David et al., 2020; Mercado-Camargo et al., 2020). From one-dimension SMILES strings to multi-dimensional representations the molecular structure is illustrated in Fig. 2. With the power of molecular representations and computational techniques, significant progress can be made in addressing this neglected tropical disease (Soiza et al., 2018).

2.2.1. *One dimensional sequence representation*

One dimensional representations typically consist of sequences of characters that correspond to the linear arrangement of atoms or functional groups within a molecule, simplifying the depiction of complex molecular structures (Tegazzini et al., 2016). SMILES strings are an example of one-dimensional representation, providing a useful way to represent the structure of small molecules in a linear, text-based form. It is an extremely short ASCII string notation that describes the atom and bond arrangement in a molecule. Therefore, it became very useful in computational processing (Robles-Loaiza et al., 2021). In anti-Leishmanial research, SMILES strings are used for high-throughput virtual screening as they enable the rapid assessment of the structural features of thousands of compounds (Soni and Pratap, 2022). Encoding molecules as SMILES strings allows for rapid prediction of toxicity and

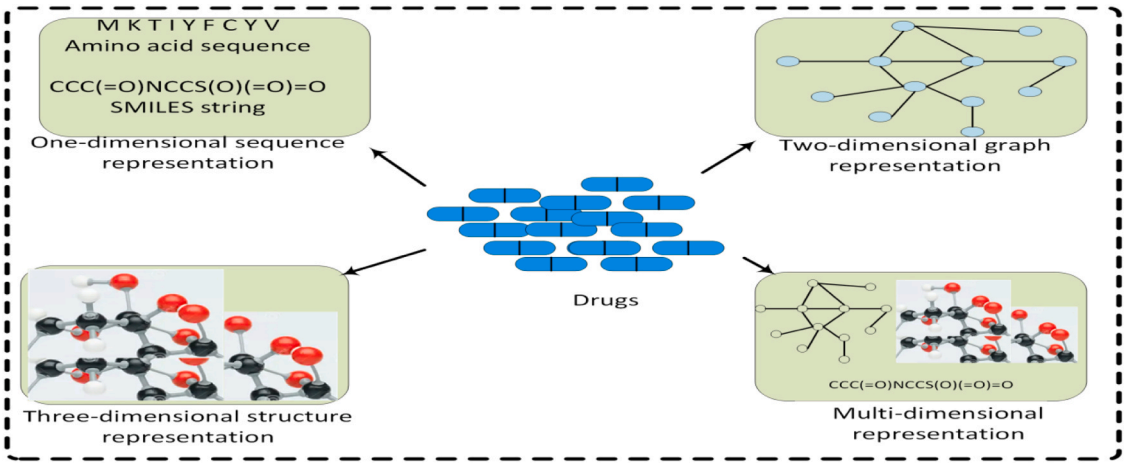


Fig. 2. Representation of drug molecules for analysis.

efficacy in drug candidates. In drug optimization and design, SMILES strings are integrated into chemoinformatics software, enabling modifications to enhance a molecules effectiveness against the *Leishmania* parasite or minimize its side effects (Cui et al., 2014). The SMILES strings of molecules serve as inputs for machine learning methods to predict various molecular properties and aid in the discovery of new anti-Leishmanial molecules (Zheng et al., 2024).

Amino acid sequence is another example of one dimensional sequence representation. Every protein consists of a linear chain of amino acids and the sequence of these amino acids determines the proteins structure and function (Brannigan and Wilkinson, 2021; Sasidharan and Saudagar, 2020). Amino acid sequences are used for *Leishmania* proteins in anti-leishmanial research to pick up possible drug targets (Duncan et al., 2011). For instance, the Leishmania genomics sequence has revealed a number of MAPK genes and they are homologous across species of the parasite. The identified sequences can be used to model protein's structure and screen for potential inhibitors. Such natural compounds as chrysin and genistein have been identified as potential inhibitors of *Leishmania* MAPK4 by means of docking simulations and homology modelling. So, one dimensional representations of amino acid sequences form the foundation for the development of drugs and therapeutic proteins based on peptides that greatly help to treat Leishmaniasis (Abdelkrim et al., 2022). All kinds of one dimensional representations are shown in Fig. 3.

Categorical features of anti-Leishmanial molecules, like molecular subtype or binding site classification, etc, can be captured using one-hot encoding. In this approach, each category has been translated to a vector of binary numbers, and hence, machine learning models find a way to process or consider categorical data (Duncan et al., 2011). K-mers are short fixed-length sequences extracted from molecular sequences of anti-leishmania compounds. Such a breakdown of sequences to k-mers provides researchers with the ability to compose a profile representative of the frequencies of each k-mer inside the molecules. This is all the more useful in discovering patterns and motifs that characterize anti-leishmania compounds (Abdelkrim et al., 2022). For instance, k-mer analysis can find common sub-sequences between molecular structures that may be used to link to high anti-leishmania activity with a view to the discovery and design of new compounds.

The molecular structure of anti-leishmania molecules can also be embedded in continuous and dimension space using methods like

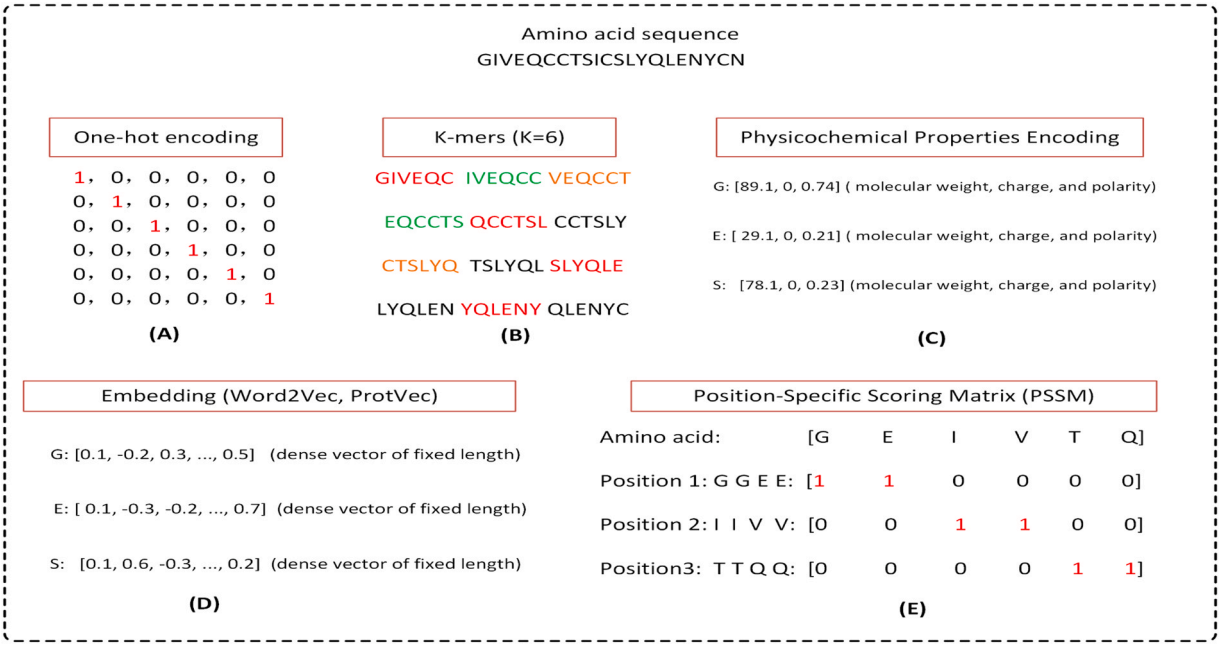


Fig. 3. One-dimensional representation of molecular data.

Word2Vec and ProtVec. Looking at k-mers as "words" and sequences of molecules as "sentences", these embedding capture the relationship that exists amongst the contextual use of molecular sequences (De Muylder et al., 2016). In this way, advanced machine learning models can be applied to classify and predict the activity of anti-leishmania molecules. For example, embedding techniques can be used to provide molecular features that are crucial for the activity of anti-leishmania compounds and thus enable techniques for the rational design of new therapeutic agents.

The PSSMs (Position Specific Scoring Matrix) are to be used to analyse the evolutionary conservation of molecular features involved in anti-leishmania compounds. Multiple sequence alignments of molecular structures give PSSMs, and thus, a matrix is obtained where each row corresponds to a position in the molecule, and every column corresponds to the probability of observing a given feature, like an atom or functional group at that position (Berhe et al., 2024). These matrices can ultimately yield better knowledge of the structural and functional features of anti-leishmania compounds, indicating related conserved features necessary for their activities. PSSMs can be used for a multitude of purposes, ranging from the location of key conserved binding sites to the prediction of the functional effects of mutations in anti-leishmania molecules.

2.2.2. *Two-dimensional graph representation*

Molecules can be viewed as graphs in their 2D representation, where atoms are represented as nodes and chemical bonds as edges, providing structured view of their connectivity (Reimão et al., 2010). This graph-based representation works very well due to the connectivity and the relative arrangement of atoms holding functional groups, as well as the ring systems that are crucial for the interactions of the molecule with its biological target (Bernal and Schmidt, 2019).

2D representations are heavily used in chemoinformatics applications involving similarity searches and Structure-Activity Relationship (SAR) studies (Kim et al., 2015). Similarity searches are the comparison of structural similarity between known anti-Leishmanial compounds and new candidate molecules in which potentially active compounds are identified due to their structural resemblance to effective drugs (Zhao et al., 2023). SAR studies, on the other hand, find the linkage between the chemical structure of a molecule and its respective biological activity so that researchers can alter molecular structures to better the efficacy and diminish the toxicity.

2D representations form an input basis for machine learning techniques, specifically Graph Convolutional Networks (GCNs), Graph Neural Networks (GNNs) (Jiang et al., 2021), Hyperbolic Graph Neural Networks (HGNN) (Lau et al., 2023) and Attention based neural network (Feng et al., 2024). These models deal directly with the graph structure; therefore, they could learn features of a molecular graph representation of the local and combined global structure. These model applied to process 2D presentations of the anti-Leishmanial drug's alleged compounds to predict properties or activities or naturally interact with the biological targets. Such models exploit the connectivity information available in these 2D graphs to provide insight into more important structural features responsible for the molecule's activity against *Leishmania*. Message Passing Neural Network(MPNN) algorithms do this by passing information through the structure of the graph in order to predict molecular properties, activities, and interactions (Swanson, 2019; Jørgensen et al., 2018; Hasebe, 2021). This makes these models particularly potent in anti-Leishmanial research because they can handle complex structures that are characteristic of many anti-Leishmanial compounds and distinguish patterns and relationships that may not be readily apparent from more simple representations (Shah et al., 2023). The two-dimensional representation of molecular data is shown in Fig. 4.

Integrating molecular fingerprinting with 2D representations enhances the prediction of molecular properties by combining structural insights with encoded molecular features. While 2D representations provide a visual or graph-based depiction of atoms and bonds, molecular fingerprinting, using tools like RDKit, captures specific molecular features such as atom connectivity and substructures (Baranwal et al., 2019). This integration allows for more accurate and comprehensive predictions, as molecular fingerprints encode essential structural information that complements the detailed 2D representation, improving the effectiveness of machine learning models in predicting molecular behavior and activity against diseases.

A molecular fingerprint is a set of numerical or binary representations describing the structure of a molecule; it is essential for predicting various molecular properties and activities (Fernández-De Gortari et al., 2017). Different kinds of fingerprints are used to emphasize several aspects of molecular structure as shown in Table 1. The most commonly used is the Morgan fingerprint, also known as ECFP (Extended Connectivity Fingerprint), a representation based on the circular neighbourhood of each atom, which therefore captures the connectivity and local sub-structure of atoms (Axen et al., 2017). The MACCS (Molecular ACCess System) key set represents sub-structural features including functional groups and other common structural motifs, and is therefore very useful in virtual screening (Cereto-Massagué et al., 2015). Atom pair fingerprints encode the relationships between atom pairs, taking into account the connectivity and distances between them, and are therefore useful for capturing long-range structural information (Capecchi et al., 2020). Other types include topological torsion fingerprints, which focus on relative localization, and heliographic fingerprints, which take a more general approach by considering higher order

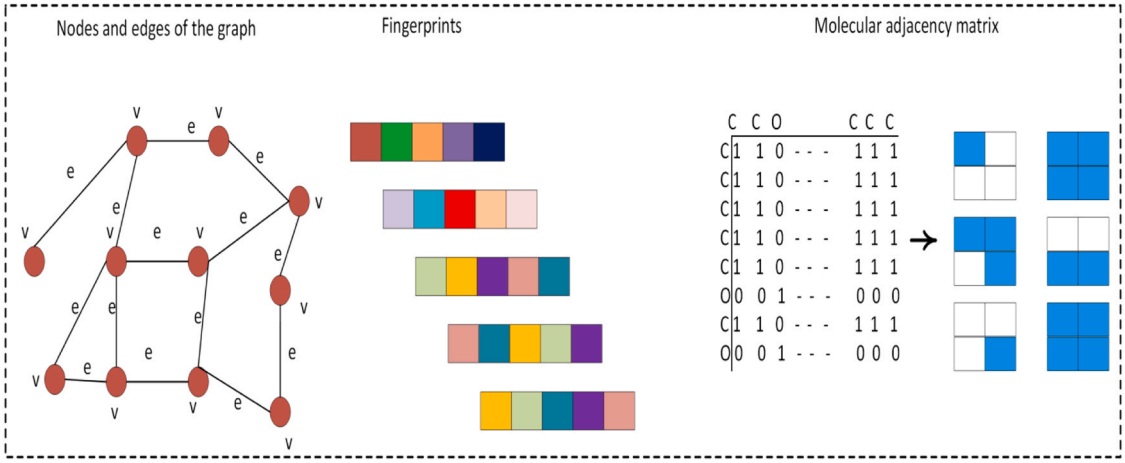


Fig. 4. Two-dimensional representation of molecular data.

Table 1
Description of various molecular fingerprints along with their usage, advantages and limitations.

Fingerprint Type	Description	Usage	Advantages	Limitations
Morgan Fingerprints (ECFP)	Circular-based fingerprints capturing atom connectivity and local sub-structures. Each atom's environment within a given radius is encoded.	Virtual screening, similarity searching, and QSAR modeling.	Highly customizable for different radii, captures local molecular structures effectively, widely supported.	Sensitive to variations in radius selection; may miss global structural features.
MACCS Keys	Binary fingerprints representing the presence or absence of 166 predefined sub-structures including functional groups and common motifs.	Structure-based virtual screening and classification tasks.	Easy to interpret due to predefined keys; widely used for quick assessments of structural similarity.	Limited by fixed key definitions; may not capture novel or complex structures.
Atom Pair Fingerprints	Encodes pairs of atoms with their chemical types and distances (in terms of bond count) between them.	Capturing long-range structural relationships, similarity searching.	Good for encoding spatial relationships and global structural information.	Sensitive to molecular size; requires more memory for large molecules.
Topological Torsion	Encodes sequences of four atoms connected by three bonds, focusing on rotational flexibility and stereo chemical properties.	identifying stereochemistry-sensitive molecular activities, conformational studies.	Captures torsional flexibility and stereo chemical information.	More computationally intensive; limited to topological (not 3D) features.
Heliographic Fingerprints	Generalized fingerprints considering atom pairs, rings, and bond types.	Broader structure-based analysis and prediction tasks.	Provides a more comprehensive representation of molecular features.	Potentially less efficient; not optimized for specific structure-property relationships.
Virtual Fingerprints (Learned Fingerprints)	Data-driven fingerprints that can be generated by machine learning models, tailored to specific tasks based on molecular datasets.	Advanced predictive modeling, dynamic interaction studies, and drug-target interaction prediction.	May be highly flexible, adaptable to complex structures including 3D conformations; can reveal hidden structure-activity relationships.	Computationally expensive; requires large training datasets and may overfit if not properly regularized.

molecular features such as atom pairs, rings and bonds (Schulz-Gasch et al., 2012). These fingerprint types vary in their application, with most choices based on the nature of what is being studied or the specificity of the biological activity. Typical applications of using machine learning models for drug discovery typically include virtual screening and target identification.

Future molecular fingerprinting techniques will employ virtual fingerprints generated by machine learning models in an attempt to represent molecular structures in a more flexible and data-driven manner. While traditional molecular fingerprinting methods rely on a predefined set of rules and algorithms to capture molecular features, virtual or learned fingerprints can be learn directly from large databases of molecular properties to reveal hidden patterns or relationships in the data. Machine learning to generate fingerprints has the advantage of being able to adjust task specificity to improve prediction accuracy by capturing more subtle, task-specific molecular features. In addition, virtual fingerprints can handle complex molecular features, such as 3D conformations and dynamic interactions, which are often difficult to describe with conventional fingerprints. Particularly in drug discovery and disease target research, traditional methods may overlook key structural details or have poor generalisation to different datasets, whereas virtual fingerprints are able to more accurately predict the molecular properties used.

2.2.3. *Three-dimensional structural representations*

Three-dimensional views of molecules take conformational flexibility and interactions into account, which is important for binding to biological targets. With this approach, researchers can get a comprehensive view of how these molecules work. In anti-leishmaniasis research, the three-dimensional structure of drug molecules becomes very important in pharmacodynamics and pharmacokinetic studies (Ugbe et al., 2022). With this knowledge, the absorption, distribution, metabolism and excretion of drugs in the body can be predicted, as well as the mechanism of action at the cellular level. Fig. 5 illustrates the three-dimensional molecular representation.

Fig. 5 shows a general deep learning workflow for analyzing drug molecules from 3D structural representations. The first step of the process is to convert the input drug molecule into a 3D density grid representation, where the spatial arrangement of the molecule is encoded in a ligand-based grid format in the form of atomic positions, charges and

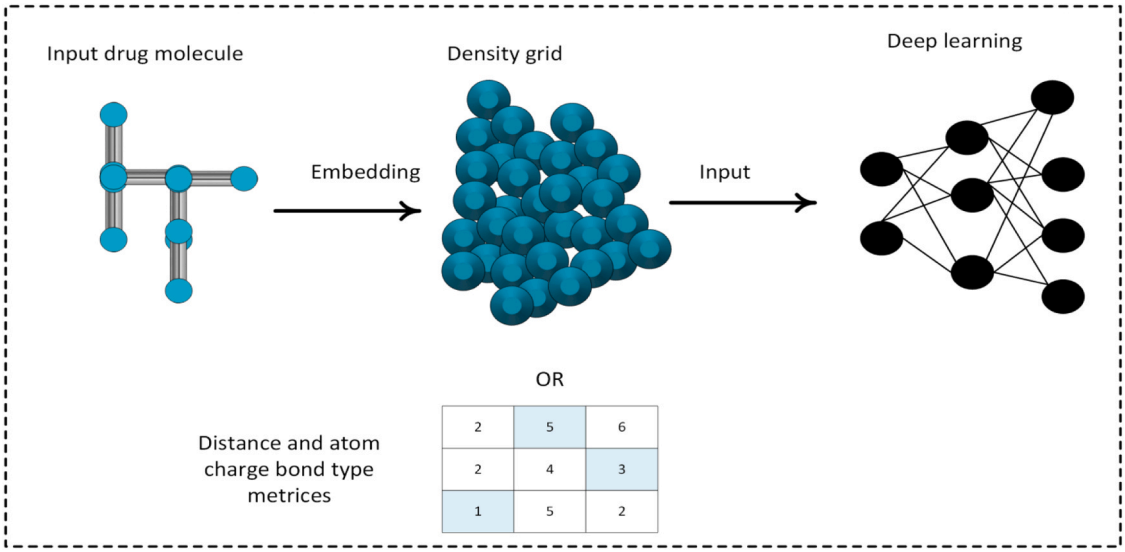


Fig. 5. Three-dimensional representation of molecular data.

bond types. It represents a three-dimensional view of the molecular structure, allowing the neural network to process its geometric and chemical properties and capture the spatial relationships between atoms. Alternatively, the basic features of the molecule can be summarized as a numerical matrix, including inter-atomic distances, atomic charges and bond types. These three-dimensional or feature-based inputs would then feed into a deep learning model that would learn from the molecule's three-dimensional structural patterns to predict its potential to become a drug candidate, such as efficacy against leishmaniasis or other biological targets.

2.2.4. Multi-dimensional hybrid representation

Multi dimensional molecular representation brings together multiple data types such as sequences, graphs and 3D structural information, and is particularly useful in the development of anti-leishmanial drugs (Prosser et al., 2020; Vlachakis and Vlamos, 2021). With this approach, it is possible to obtain a holistic view containing the intricate relationships and interactions within a molecule; this may be overlooked if a single data format is used (Rohini et al., 2019). By integrating different data modalities, this multidimensional characterisation provides a detailed understanding of molecular properties and behaviour, leading to a better understanding of their therapeutic potential and efficacy (Mazón-Ortiz et al., 2023; Abdoul-Latif et al., 2024)]. In anti-leishmaniasis research, it is possible to integrate gene sequence data of *Leishmania* enzymes, 2D graphical structures of drug compounds, and 3D structural information of target proteins and drug molecules. These integrated data have the potential to allow researchers to more closely examine how drug candidates interact with each other in terms of potency, safety, and selectivity (Blanco-González et al., 2023). Combining sequence data with 3D structural information can shed light on how mutations in enzymes affect drug binding and resistance mechanisms. As shown in Fig. 6, this multivariate strategy in machine learning increases the dimensionality of prediction and reveals important features.

Different types of data can be used to derive a variety of features that can be utilized by machine learning models to train more accurate and robust models to predict the efficacy, safety, and possible off-target effects of molecules. Importantly, if models are trained on the basis of such multidimensional data, it is possible to predict not only the potential efficacy of compounds against leishmaniasis, but also their potential toxicity and side effects in humans. This kind of information is absolutely invaluable for the screening phase of large compound libraries and for selecting the most promising candidates for further development. As a result, researchers seeking multidimensional characterization can design more effective and specific anti-leishmanial therapies. This approach both accelerates the drug discovery process and limits adverse side effects, thereby increasing the success rate of viable anti-leishmanial therapeutic drug development. The comparison of various dimensional molecular representations is shown in Table 2.

2.3. Feature extraction

Extracting features from molecules after their representation is a key step in many computational applications. Different molecular representation formats, such as 1D, 2D, 3D, and multi-dimensional, require specific techniques to capture their unique features. In one study, statistical and machine learning techniques were used to select important features like patient demographics, lesions, and environmental factors, which improved the accuracy and efficiency of diagnostic methods for cutaneous Leishmaniasis. Feature selection identified the most influential variables, leading to more focused approaches that enhanced diagnostic accuracy and patient care (Rahi, 2013). Another study used clinical features such as lesion size, location, and morphology, combined with PCR techniques, to extract relevant data. This helped differentiate between Leishmania species, improving diagnosis and enabling targeted treatments (Alkalah, 2016). Some common techniques for feature extraction from different molecular representations are shown in Table 3.

Various techniques are employed to extract meaningful features from molecular data across different dimensions. SMILES tokenization, one-hot encoding, and bag-of-substructures are commonly used to process linear sequences, capturing atomic and sub-structural information (McGibbon et al., 2024; David et al., 2020). For graph-based molecular data, techniques like Graph Convolutional Networks (GCNs), molecular fingerprinting, and adjacency matrices are utilized to analyze connectivity and both local and global features of molecules (Nguyen and Nguyen-Xuan, 2020). Spatial molecular data is processed using 3D Convolutional Networks (3D CNNs), distance matrices, and molecular descriptors, which exploit geometric and conformational information. For comprehensive molecular analysis, advanced methods such as feature fusion, multi-modal learning, and attention mechanisms integrate multiple data types to provide a holistic understanding of molecular properties. These techniques are critical for applications such as

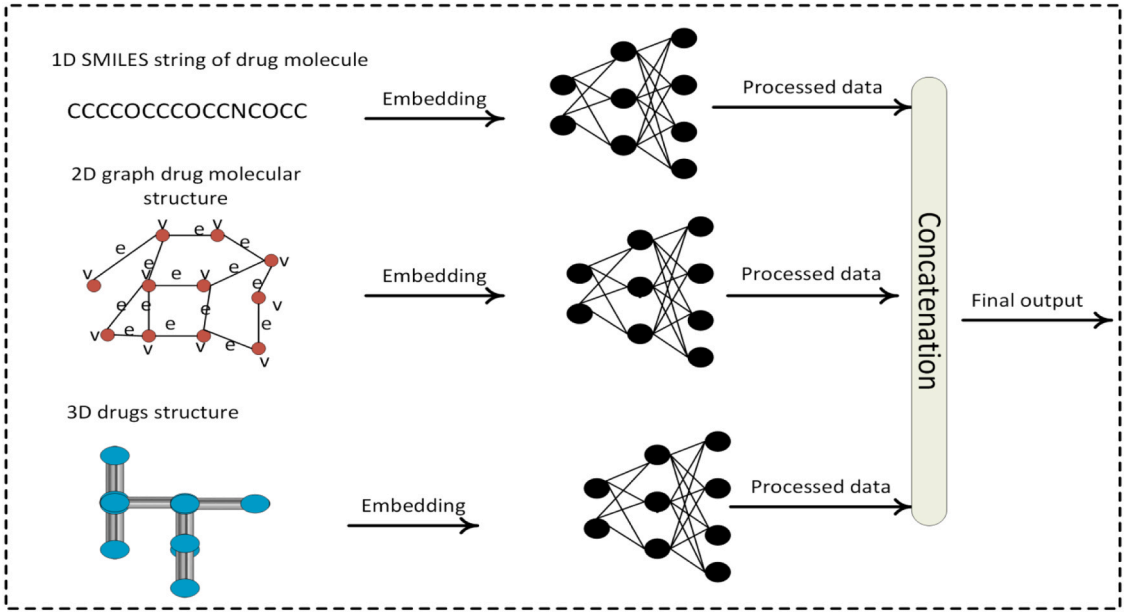


Fig. 6. Multi-dimensional input of molecular data into the machine learning models.

Table 2
Comparison of molecular data representation.

Representation	Description	Key Features	Advantages	Limitations
1D Representation	Linear representation of molecules (e.g., SMILES strings or genetic sequences).	Encodes molecular structure or sequences in a simple, linear format.	Simple, compact, easy to process; suitable for sequence-based analysis.	Loses spatial and structural information; limited insight into interactions.
2D Representation	Molecular graphs showing atoms as nodes and bonds as edges.	Captures connectivity and relationships between atoms and bonds in a flat, 2D plane.	Visualizes molecular structure; suitable for identifying functional groups and bond networks.	Lacks 3D spatial data; cannot capture stereochemistry or spatial interactions.
3D Representation	Three-dimensional structural data (e.g., density grids, atomic coordinates).	Encodes spatial arrangement of atoms, bond angles, and molecular geometry.	Provides detailed insights into molecular geometry and interactions, including docking with target sites.	Computationally intensive; may require complex preprocessing for machine learning models.
Multi-dimensional Representation	Integrates multiple data types (e.g., sequences, 2D graphs, 3D structures, and additional properties).	Combines data from multiple modalities, such as genetic sequences, molecular graphs, and 3D structures.	Captures a holistic view of molecular properties and interactions; enhances predictions in ML applications.	Data integration can be complex; requires advanced algorithms for processing and interpretation.

Tablee 3
Feature extraction techniques from molecular data Dimension data.

Data dimension	Techniques	Description
1D	SMILES Tokenization, One-Hot Encoding, Bag-of-Substructures	Processes linear sequences to capture atomic and sub-structural information
2D	Graph Convolutional Networks (GCNs), Fingerprinting, Adjacency Matrix	Utilises graph structures to capture connectivity and local/global molecular features
3D	3D Convolutional Networks (3D CNNs), Distance Matrix, Molecular Descriptors	Exploits spatial coordinates to capture geometric and conformational information
Multi-dimensional	Feature Fusion, Multi-Modal Learning, Attention Mechanisms	Integrates various data types to capture comprehensive molecular characteristics

drug discovery and diagnostics, enabling the extraction of specific and relevant molecular features.

2.4. Feature selection

Feature selection plays a crucial role in computational chemistry and drug discovery applications, especially in QSAR model development. When thousands of descriptors are used to describe a molecule, selecting the most relevant and informative subset is essential for building robust and interpretable models (Goodarzi et al., 2012). In the context of Leishmaniasis, Ramezankhani et al (Ramezankhani et al., 2018). proposed a decision tree algorithm to select variables for predicting the incidence of cutaneous Leishmaniasis. This study used environmental and topographic factors, such as altitude, temperature, and humidity, as input variables to predict regions where outbreaks are most likely to occur. The decision tree effectively identified key predictive variables, enabling focused public health interventions. The research also highlighted how selecting the right features can optimize disease prediction for Leishmaniasis in specific regions. Proper feature selection enhances model performance, stability, and generalization, while reducing overfitting and computational complexity (Ponzoni et al., 2017). Fig. 7 illustrates the feature selection process.

Various feature selection methods have been applied in QSAR studies, including filter methods, e.g., information gain, correlation-based filter, wrapper methods where subsets are evaluated regarding the model performance, embedded methods, which are inherently part of the modelling algorithm itself and hybrid approaches (Goodarzi et al., 2012; Guha and Willighagen, 2020). Commonly used algorithms for feature selection in QSAR are recursive feature elimination, random forests, and regularisation techniques such as LASSO and elastic net (Moriwaki et al., 2018).

2.5. Predictive models

Leishmaniasis remains a major public health issue, and researchers are exploring many treatment options. Croft and Coombs et al (Croft and Coombs, 2003). provide an overview of chemotherapy drugs for Leishmaniasis, highlighting miltefosine as the primary drug currently used, and discuss recent progress in discovering new treatments. They stress the need for better therapies, as existing treatments have many limitations and side effects. Sundar and Chatterjee (Sundar and Chatterjee, 2006), reviewed treatment options for visceral Leishmaniasis, focusing

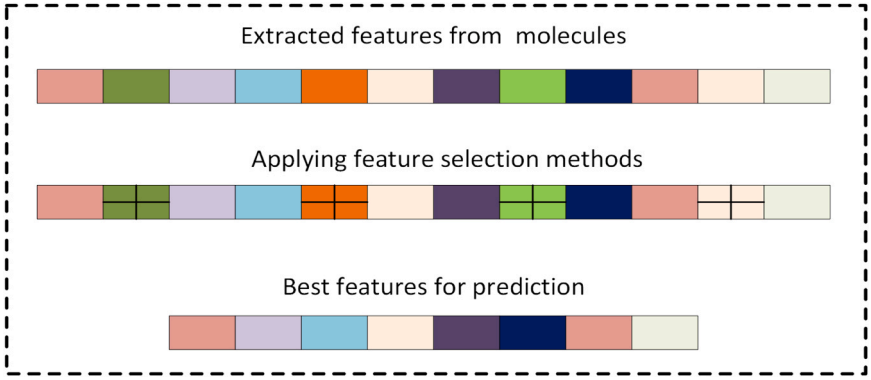


Fig. 7. Illustration of the feature selection process.

on the effectiveness of Amphotericin B, while also pointing out the urgent need for safer and more affordable alternatives.

One key study by Jha et al. (2005) shows that a single-dose liposomal version of amphotericin B is as effective in treating visceral Leishmaniasis as longer treatments, reducing the treatment time and improving patient compliance. This marks a significant improvement over earlier treatment methods for this debilitating disease. In contrast, Lira et al. (1999) discuss treatment failures in India due to resistance of Leishmania donovani isolates to antimony, highlighting the need for alternative treatment strategies to combat drug resistance.

Jebbari and Davidson (1998) discuss the introduction of miltefosine as the first oral treatment for visceral Leishmaniasis, emphasizing the importance of ensuring access to this life-saving breakthrough in endemic regions. In another study, Croft and Engel (2006) explain how miltefosine, initially developed for cancer treatment, later showed its potential as an anti-Leishmanial drug. Additionally, the study by Sundar and Chatterjee (2006) investigates sitamaquine as an oral alternative for visceral Leishmaniasis, noting its promise due to its effectiveness and the convenience of oral administration.

Beyond traditional drug discovery, machine learning and computational methods have significantly accelerated Leishmaniasis research, particularly in drug discovery, diagnostics, and ecological modeling. A machine learning-based chemoinformatics model was developed to predict pyruvate kinase inhibitors in Mexican gooseberry mushrooms. The accuracy, precision and recall of the model were 85 %, 82 % and 80 %, respectively. Chemoinformatics modeling can effectively screen inhibitors in drug discovery, thereby reducing the need for extensive experimental testing (Rahi, 2013).

Machine learning methods have been applied to disease diagnosis, primarily through image analysis, to improve the detection of

Table 4
Application of Machine Learning: Strengths and Weaknesses in Leishmaniasis.

ML Method	Strengths	Weaknesses	Applications in Leishmaniasis	Example
Random Forest (RF)	Handles large, high-dimensional data.Robust to overfitting.Good for feature selection.	Can be computationally expensive. Less interpretable.	Classification of compounds as active/inactive.Feature selection for drug screening.	RF models for virtual screening of antileishmanial compounds.
Support Vector Machine (SVM)	Effective in high-dimensional spaces.Can model complex relationships.Robust to overfitting.	Requires careful tuning. Limited interpretability.	Classification of antileishmanial activity. Prediction of bioactivity.	SVM models for classifying compounds with leishmanicidal activity.
K-Nearest Neighbors (KNN)	Classifies data based on the proximity of data points to others.	Sensitive to noisy data and irrelevant features.	Simple to implement, effective with smaller datasets, interpretable.	Classifying compounds based on chemical structures for potential Leishmania drug discovery.
Convolutional Neural Networks (CNNs)	Used for analyzing images or molecular structures, recognizing patterns in complex data.	Requires large datasets and high computational power.	Powerful for modeling complex, non-linear relationships, great for high-dimensional data.	Predicting the biological activity of compounds by analyzing molecular structures in drug discovery for Leishmania.
Recurrent Neural Networks (RNNs)	Models sequential data, capturing temporal or sequential dependencies in data.	Can be difficult to train due to vanishing gradient problem, especially for long sequences.	Effective for modeling sequences, good for predicting molecular interactions and drug resistance.	Predicting molecular interactions or drug resistance evolution in Leishmania infections.
Autoencoders	Unsupervised learning technique for dimensionality reduction, preserving important features.	Can struggle with very noisy or unstructured data.	Useful for feature extraction and reducing dimensionality, can improve model performance.	Reducing the dimensionality of molecular datasets and enhancing drug discovery processes for Leishmania.
Non-negative Matrix Factorization (NMF)	Extracts latent factors from datasets, particularly used in collaborative filtering tasks.	Requires careful initialization and can be sensitive to noise.	Reveals hidden patterns in data, effective for matrix-based data like compound-target interaction data.	Identifying novel anti-Leishmania compounds by extracting latent factors from biological data.
Singular Value Decomposition (SVD)	Decomposes matrices into simpler components to reveal underlying patterns.	Can be computationally expensive for very large datasets.	Reduces dimensionality, allows for better data interpretation, effective for large-scale biological data.	Analyzing compound-protein interaction data and predicting drug candidates for Leishmaniasis.
Graph Convolutional Networks (GCNs)	Uses graph-based representations (atoms as nodes, bonds as edges) to model molecular structures.	Computationally intensive, especially for large graphs.	Effective at capturing molecular relationships, directly models chemical structures.	Identifying potential anti-Leishmania compounds by modeling molecular interactions as graphs.
Graph Neural Networks (GNNs)	Extends GCNs to more complex graph-based data, predicting molecular properties and drug-target interactions.	High computational demand, may require large amounts of training data.	More flexible and accurate than traditional GCNs, better at predicting molecular properties and toxicity.	Predicting drug-target interactions, molecular properties, and toxicity for anti-Leishmania drug discovery.
Hyperbolic Graph Convolutional Networks (Hyperbolic GCN)	Uses hyperbolic space to represent hierarchical data and capture richer structural relationships.	Hyperbolic space models are more complex and can be difficult to optimize.	Better at modeling hierarchical structures, provides deeper insights into complex graph data.	Modeling hierarchical chemical or biological interactions for more accurate prediction of Leishmania drug candidates.
Message Passing Neural Networks (MPNN)	Uses message passing between nodes in a graph to update node representations and predict molecular properties or drug interactions.	Requires significant computational resources and tuning.	Very effective in modeling complex molecular interactions, adaptable for various drug discovery tasks.	Predicting drug-target interactions and molecular properties, useful in optimizing compounds for anti-Leishmania drug development.
Boosting (e.g., AdaBoost, Gradient Boosting)	Combines multiple weak learners into a strong model to improve prediction accuracy.	Sensitive to noisy data, may overfit if not tuned properly	Enhances the performance of weak models, robust to overfitting if used properly	Predicting the efficacy of anti-Leishmania drugs by combining weaker models into a strong prediction system.
Transfer Learning	Leverages a pre-trained model on one task and adapts it to a new but related task, often with less data.	May require fine-tuning, model selection can be challenging.	Allows faster model training on smaller datasets, can leverage existing knowledge, reduces training time	Can be used in Leishmania drug discovery by adapting models trained on other diseases or tasks (e.g., molecular activity prediction) to improve prediction accuracy with limited data.

leishmaniasis in microscopy images. In the study "Machine Learning Based Leishmaniasis Detection System", the model demonstrated 68 % accuracy, 60 % specificity and 51 % sensitivity, proving its robustness and reliability in disease diagnosis (Zare et al., 2022).

Ecological modeling has been further refined using machine learning, especially in terms of GIS-based prediction of the distribution of the Golestan skin leishmaniasis vector, the sand fly (Phlebotomus papatasi). Thus, these methods helped to indicate high priority areas that should be targeted for intervention. The model performed well with an AUC score of 0.87, precision of 86 % and recall of 84 %. This is important for focusing efforts on vector control to avoid disease transmission (Mollalo et al., 2018).

In addition, meta-analysis with machine learning has been applied to study the transcriptional response of immune cells against leishmaniasis infection. It identifies gene expression patterns to better understand the immune mechanisms that can be targeted for therapy ("Meta-analysis and Machine Learning in the Immune Response to Leishmaniasis"). The developed model has an AUC of 0.89 on unseen data, a precision of 88 %, and a recall of 87 %, thus performing well in predicting transcriptional responses associated with activated immune cells. These results shed further light on the immune mechanisms of leishmaniasis and may guide future therapeutic strategies (Rezaei et al., 2024). Based on the above studies, the following review will focus on how different machine learning methods can be applied to predict the anti-leishmaniasis molecules listed in Table 4.

2.5.1. Classical machine learning methods in Leishmania disease

In general, ML methods have been used in different studies related to drug discovery and prediction of disease (Zare et al., 2022; Huang et al., 2022). RF, KNN, and SVM are some of the most commonly used classification techniques for identifying molecules for potential activity against leishmaniasis. For example, RF models have been developed to predict the biological activity of new compounds against leishmaniasis parasites by analyzing chemical features and optimization of hyperparameters (Singh and Kumar, 2020). KNN has also been applied to classify compounds, which have structural similarities with already active compounds, and thus this approach allows the identification of new compounds as potential drugs. Similarly, SVM has been used to distinguish active compounds from inactive compounds based on molecular structure an important procedure for screening large libraries of compounds to identify potential therapeutic candidates (Bamorovat et al., 2021).

In the future, improvements in the application of classical ML methods in drug discovery for leishmaniasis can be accomplished with more and larger different datasets. Future research should be directed toward developing more accurate and interpretable methods by integrating multi-omics data into predictive models. Besides, combining classical Machine learning methods with other advanced techniques like deep learning could enhance their performance. The focus could also be on developing hybrid models that integrate classical ML with genetic data or biochemical from Leishmania parasites, thereby improving the prediction of drug efficacy and resistance mechanisms. Furthermore, in future research, the application of classical ML to personalized medicine, i.e., tailoring predictions to the patient's individual situation, could be explored.

2.5.2. Deep learning approaches in Leishmania disease

Deep learning methods have recently been promising in several areas of biomedical research, including drug discovery for leishmaniasis (Oualha et al., 2024). CNNs have been used to analyse molecular structures, to predict the bioactivity of compounds against the leishmaniasis parasite, and to identify drug candidates (Ugbe et al., 2022). These applications include the application of RNNs to model the evolution of drug resistance, thus providing insight into the time-dependence of the dynamics of parasite-drug interactions (Winkler, 2021). Similarly, autocoders have been applied to the downscaling of molecular datasets with the aim of capturing important features and patterns that can inspire the drug discovery process. As a result, these DL models have revolutionized the identification of potential therapies for leishmaniasis, as evidenced by their big data processing power and ability to capture non-linear relationships.

Future research directions in Leishmaniasis drug discovery may be to harness the power of deep learning to more accurately understand the complex molecular interactions and processes underlying drug resistance. One such direction would be to combine multi-modal data such as genomics, transcriptomic and proteomics with deep learning models to improve the prediction of compound efficacy and resistance. Reinforcement learning could also be explored to optimize drug discovery pipelines by adapting to new data in real time and improving model accuracy. Another promising area is the use of migration learning to apply pre-trained models for other diseases to the drug discovery task for Leishmaniasis, thereby reducing the need for large amounts of data. In addition, the application of deep learning models to predict the effect of compounds on immune evasion strategies in Leishmaniasis could provide new avenues for therapeutic strategies.

2.5.3. Matrix factorization techniques in Leishmania disease

Matrix factorization methods i.e. such as Non-negative Matrix Factorization (NMF) and Singular Value Decomposition (SVD) have been widely used in drug discovery, thus revealing large-scale biological and chemical datasets with hidden patterns (Moon and Lee, 2021). On the other hand, SVD can be used to analyse compound-protein interaction data, thereby predicting drug candidates from the interaction of small molecules with *Leishmania* proteins. Both of the discussed matrix factorization methods are suitable for the case of very sparse data. These methods can reveal hidden structures in intricate biological processes and thus will certainly be of great use in drug discovery for Leishmaniasis.

In addition, future work may focus on enhancing matrix factorization techniques, e.g. to further improve the performance of in-silico drug target prediction for improved discovery of anti-leishmaniasis compounds. This could be further extended to the prediction of drug effects on leishmaniasis at the systems biology level, integrating the latest knowledge of complex biology represented by high-throughput IoT data (e.g., gene expression or protein interaction data). In addition, matrix factorization can be combined with other advanced methods, such as graph-based learning or deep learning approaches, to improve the accuracy and generalizability of predictions. Similarly, the development of hybrid models combining matrix factorization with molecular docking simulations provides a holistic concept for predicting the binding affinity of compounds to leishmaniasis targets, which will undoubtedly improve drug discovery.

2.5.4. Graph-based learning methods in Leishmania disease

Graph-based learning methods include graph convolution networks and graph neural networks, both of which represent molecular structures as graphs with atoms as nodes and bonds as edges (Wang et al., 2024). In graph convolutional networks, the prediction of a compound's biological activity is achieved by analyzing the molecular graph; this will always conform to the relationships of atoms within the molecule. GNN further develops this methodology with greater flexibility and higher accuracy in predicting drug-target interactions and toxicity. These methods are ideally suited to the drug discovery process for leishmaniasis, where the molecular structure and its interactions are so complex that a detailed understanding is required to identify effective treatments.

In the future, these graph-based learning methods will have tremendous scope for expansion in Leishmaniasis drug discovery as multi-scale data are integrated into the model. The use of hyperbolic graph networks can capture hierarchical relationships to better model leishmaniasis-host interactions and provide insight into drug-host-pathogen dynamics. Adding temporal data can further model the

progression of Leishmania protozoa infections or the emergence of drug-resistant strains. Predictive capabilities will be further improved when AI methods for graph-based modeling are developed, allowing for the interpretation of the contribution of each specific molecular feature or interaction to the prediction, thus enabling safer and more effective design of anti-leishmanial drugs.

2.5.5. Ensemble learning techniques in Leishmania disease

Ensemble learning methods, including Boosting-based methods such as AdaBoost and Gradient Boosting, as well as Bag Learning methods such as Random Forest, have been used in the drug discovery process for Leishmaniasis (Singh and Kumar, 2020). They aggregate several weaker models into one strong model, resulting in high overall performance, which is popular in complex prediction efforts such as prediction of the anti-leishmanial compounds. Ensemble methods such as augmentation have been tried in order to combine multiple classifiers to improve prediction accuracy in terms of efficacy and safety. In addition, some bagging techniques such as Random Forest have been used to reduce the variance of the predictions and thus make the model more robust (Shah et al., 2022). These ensemble methods are especially valuable in situations where data is noisy or where the generalization ability of the model is important.

In the near future, it is likely that advanced ensemble machine learning or deep learning techniques will collaborate to come up with better Leishmaniasis drug discovery programs with robust predictions. A potential direction is the use of stacking, where different machine learning models are combined in a hierarchical manner to refine predictions. Additionally, research could focus on adapting ensemble methods for the specific challenges of *Leishmania* such as predicting the effects of drugs on different Leishmania species or accounting for the diversity in host responses to treatment. Furthermore, ensemble learning could be combined with reinforcement learning to create adaptive models that continuously improve their predictions based on incoming experimental data.

2.5.6. Transfer learning in Leishmania disease

Recently, transfer learning has begun to take root in the field of drug discovery, including leishmaniasis, where researchers can utilize limited data to apply models pre-trained on large datasets to new tasks (Cai et al., 2020; Buterez et al., 2024). This approach proved particularly valuable when there was insufficient labeled data to train from scratch. For example, pre-trained models from other parasitic disease datasets (e.g., malaria) have been used to predict anti-leishmaniasis drug candidates. The process of transfer learning can be used to utilize knowledge gained from larger datasets, thereby improving the efficiency and accuracy of predictions in the presence of limited data. This would greatly reduce the need for large datasets, thereby speeding up the process of anti leishmanial drug discovery.

In the future, transfer learning related to Leishmaniasis drug discovery will be oriented towards integrating diverse data from multiple sources, such as proteomics, genomic and clinical data in order to further improve the transferability of pre-trained models. These researchers may investigate the use of domain-specific data to fine-tune pre-trained models to improve the therapeutic efficacy of leishmaniasis. Some other promising research direction is to apply the principles of transfer learning to personalized medicine, i.e., to fine-tune models trained on large datasets of Leishmaniasis for a specific patient's condition to enable personalized treatment strategies. Likewise, the use of transfer learning in combination with other modern techniques such as reinforcement learning or ensemble methods can also improve performance, leading to improved prediction accuracy and drug efficacy against leishmaniasis infections.

3. Evaluation metrics and cross validation

In machine learning, evaluation metrics and validation strategies are considered as two important aspects to assess the performance and generalization of machine learning models (Cabitza et al., 2021). Metrics such as accuracy, precision, recall, F1-score and AUC-ROC are commonly used to quantify the performance of predicting any outcome, such as identifying compounds with potential anti-leishmanial activity or diagnosing a disease. Validation methods such as cross-validation, holdout validation and early stopping validation measure the robustness of the model to high performance on any unseen data, thus avoiding overfitting. All of these have a very important place in any research related to leishmaniasis as they offer the possibility of improving the predictive performance and reliability of models in the search for effective therapeutic and diagnostic approaches against this neglected tropical disease. As several methods for predicting leishmaniasis are discussed in detail in Section 2, these methods can be evaluated and validated using various metrics, as shown in Table 5.

4. Discussion

Leishmaniasis is a chronic disease caused by the *Leishmania* parasite that affects many people in the tropics and subtropics. Since this infection is chronic, it has become very important to find various ways of combating it at the individual and group level (Roatt et al., 2014; Abadías-Granado et al., 2021; Saini et al., 2022). A number of methods have been used to predict treatments that may work against Leishmania parasites and then observe the outcome of the disease (Ruggeri et al., 2014; Colby et al., 2017; Andersen and Winter, 2019). These methods include traditional and modern approaches to develop new molecules with high anti-leishmanial activity that can reach therapeutic levels to reduce the disease burden. In this direction, various computational methods, *in vitro*, *in vivo*, molecular docking and other predictive techniques are expected to provide effective solutions to combat leishmaniasis or other diseases (Cunningham and Fairlamb, 1995; Beverley et al., 1986; Ye et al., 2023).

In vitro bio-assays provide primary information on efficacy; *in vitro* compound assays on cultured parasites under controlled conditions provide additional information on how the molecules act directly on leishmaniasis cells. On the other hand, complementary in vivo

Table 5
Evaluation metrics and validation strategies of prediction methods.

Model type	Evaluation metrics	Validation Strategies
Classical Machine Learning	Accuracy Precision, Recall, F1-Score AUC-ROC Confusion Matrix	K-fold Cross-validation Stratified Cross-validation
Deep Learning	Loss Function (Cross-entropy loss, Mean Squared Error), Precision, Recall, F1-Score, AUC-ROC Confusion Matrix	Holdout Validation, K-fold Cross-validation Early Stopping
Graph Learning	Graph-Based Metrics (Average shortest path, Clustering coefficient, Degree centrality), Node and Edge Prediction Metrics (MAE, RMSE) AUC-ROC	Graph Cross-validation (no data leakage between molecules), Node-wise Cross-validation
Matrix Factorization	Mean Squared Error (MSE) Root Mean Squared Error (RMSE) AUC-ROC, Precision, Recall, F1-Score	K-fold Cross validation, Reconstruction Error (how well matrix is approximated)
Ensemble Learning	Accuracy, Precision, Recall, F1-Score AUC-ROC	Bagging and Boosting (internal handling of data splitting), Cross-validation to assess stability and robustness
Transfer Learning	Accuracy, Precision, Recall, F1-Score AUC-ROC, Loss Function (Cross-entropy, MSE)	Fine-tuning and Transferability Check, Cross-dataset Validation (testing how well the model generalizes to other datasets)

evaluations of these models allow estimation of therapeutic efficacy, pharmacodynamic response, and toxicity profiles based on living organisms, thus providing data on primary bioavailability, systemic/toxicity, and therapeutic framework. Despite these advantages, traditional experimental approaches still face many significant challenges that computational methods are designed to address: in vitro and in vivo experiments are resource-intensive, time-consuming and costly for large-scale screening of chemical libraries; the reproducibility of experiments can be affected by changes in biological systems and environmental factors; and thirdly, ethical considerations prohibit the conduct of large numbers of animal experiments, especially when there are alternative methods are available.

ML and DL methods overcome these limitations by utilizing large datasets of available biological, chemical, and pharmacological information to build predictive models that can simulate experimental outcomes. For example, ML algorithms can predict the activity of compounds against specific targets in leishmaniasis, classify molecules based on potential toxicity, or model resistance mechanisms. DL architectures, including CNNs, MPNN and GNNs, can capture complex molecular structures and relationships between atoms and bonds to improve prediction accuracy. These advances in computational technology reduce reliance on experimental screening, accelerate the process of drug discovery, and pave the way for more efficient validation efforts by experimentalists, thereby accelerating the pace of effective anti-leishmanial therapies and reducing resource and ethical concerns.

This review proposes a holistic machine learning approach for predicting potent anti-leishmanial activity. The pipeline proposed in this paper, in combination with multiple machine learning techniques, can provide an effective method for analyzing and predicting molecular structure and biological activity for the development of potent anti-leishmanial disease molecules. The study aims to find compounds that could be potential candidates for leishmaniasis drug development. Therefore, this study reviews the different methods applied at this pipeline level and critically reviews the use, advantages, limitations and challenges of these methods.

Machine Learning (ML) and Deep Learning (DL) can be of great benefit for the discovery and development of molecules for leishmaniasis. By using large datasets, these methods can determine which chemical structures and biological activities are complementary, leading to the rapid discovery of good therapeutic candidates. In this regard, ML and DL techniques such as CNNs and GNNs are particularly well suited to analyze complex data from molecular databases, including data based on molecular fingerprints and SMILES representations. These models are able to learn complex features from chemical and biological data related to compound efficacy, toxicity, and their interactions with target proteins; thus, such models can reduce time-consuming and expensive experimental screening in the early stages, thereby increasing the speed of the drug discovery process. In addition, ML/DL models can be used to predict the efficacy of existing leishmaniasis drugs as well as to discover new leishmaniasis drugs.

There are some challenges in the discovery of anti-leishmanial drugs using the ML/DL approach. The main issues are two-fold: availability and quality of datasets, especially for diseases like leishmaniasis, where comprehensive datasets including both active and inactive compounds are almost non-existent. In addition, the imbalance between a few target compounds and a large number of inactive compounds may bias the predictions towards the major classes. On top of that, most state-of-the-art models, especially DL methods, are not interpretable, which is an important feature for understanding the rationale behind predictions, especially in biomedical applications. High-dimensional biological data poses another challenge: computational complexity, as the training and deployment of models requires significant computational resources. All of these challenges point to the need for sustained collaboration between computational approaches and experimental validation to overcome limitations and advance the discovery of drugs related to leishmaniasis.

5. Conclusion

This review describes a high-throughput machine learning computational pipeline for the prediction of potent anti-leishmanial molecules. It can easily process high-dimensional data to predict molecular activities against Leishmania parasites well. This will not only help in identifying promising drug candidates, but also improve the efficiency of the drug discovery process against neglected tropical diseases such as leishmaniasis. In addition to anti-Leishmania protozoa activity prediction, the proposed machine learning computational pipeline further extends its use in many other directions; it can be easily extended to a variety of other applications related to molecular characterization, drug toxicity, and other pharmacological properties. Extending this computational pipeline with more functionality and larger datasets may bring additional value to determining the relationship between structure and biological activity, and the resulting benefits may be applicable to all aspects of the drug development process.

Funding

The authors extend their appreciation to the Deanship of Research and Graduate studies at King Khalid University for funding this work through Large Research project under grant number RGP. 2/8/46. Sponsor has no role for design and publication.

CRediT authorship contribution statement

Ali Shah Hayat: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Ansari Mohammad Yousuf:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Methodology, Formal analysis. **Yasmin Sabina:** Writing – review & editing, Visualization, Methodology, Funding acquisition, Formal analysis.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) did not use any AI Tool Service.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors extend their appreciation to the Deanship of Research and Graduate studies at King Khalid University for funding this work through Large Research project under grant number RGP. 2/8/46. Sponsor has no role for design and publication.

References

Abadías-Granado, I., Diago, A., Cerro, P.A., Palma-Ruiz, A.M., Gilaberte, Y., 2021. Cutaneous and Mucocutaneous Leishmaniasis. Actas Dermosifiliogr. 112 (7), 601–618. https://doi.org/10.1016/j.ad.2021.02.008.

Abdelkrim, Y.Z., et al., 2022. Enzymatic and molecular characterization of anti-leishmania molecules that differently target leishmania and mammalian eIF4A proteins, LieIF4A and eIF4AMus. Molecules 27 (18). https://doi.org/10.3390/molecules27185890.

Abdoul-Latif, F.M., et al., 2024. Overview of research on Leishmaniasis in Africa: current status, diagnosis, therapeutics, and recent advances using by-products of the sargassaceae family. Pharmaceuticals 17 (4). https://doi.org/10.3390/ph17040523.

Agu, P.C., et al., 2023. Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. Sci. Rep. 13 (1), 1–18. https://doi.org/10.1038/s41598-023-40160-2.

Alkalah, C., 2016. Strategi Promosi Pengurus POKJAR Mandiri Pemalang dalam Menjaring Calon Mahasiswa Baru Universitas Terbuka di Era Society 5.0. Ganaya: Jurnal Ilmu Sosial Dan Humaniora 19 (5), 1–23.
Alvar, J., et al., 2012. Leishmaniasis worldwide and global estimates of its incidence. PLoS One 7 (5). https://doi.org/10.1371/journal.pone.0035671.
Andersen, M.L., Winter, L.M.F., 2019. Animal models in biological and biomedical research – experimental and ethical concerns. An. Acad. Bras. Cienc. 91, 1–14. https://doi.org/10.1590/0001-3765201720170238.
P. Anuntasomboon *et al.*, Genome alteration of Leishmania orientalis under Amphotericin B inhibiting conditions, pp. 1–28, 2024, doi: 10.1371/journal.pntd.0012716.
Axen, S.D., Huang, X.P., Cáceres, E.L., Gendelev, L., Roth, B.L., Keiser, M.J., 2017. A simple representation of three-dimensional molecular structure. J. Med. Chem. 60 (17), 7393–7409. https://doi.org/10.1021/acs.jmedchem.7b00696.
Bamorovat, M., et al., 2021. A novel diagnostic and prognostic approach for unresponsive patients with anthroponotic cutaneous leishmaniasis using artificial neural networks (May). PLoS One 16 (5), 1–17. https://doi.org/10.1371/journal.pone.0250904.
Baranwal, M., Magner, A., Elvati, P., Saldinger, J., Violi, A., Hero, A.O., 2019. A deep learning architecture for metabolic pathway prediction. no. 2010 Bioinformatics 1–7. https://doi.org/10.1093/bioinformatics/btz954.
Berhe, H., Kumar Cinthakunta Sridhar, M., Zerihun, M., Qvit, N., 2024. The potential use of peptides in the fight against chagas disease and Leishmaniasis. Pharmaceutics 16 (2). https://doi.org/10.3390/pharmaceutics16020227.
Bernal, F.A., Schmidt, T.J., 2019. A comprehensive QSAR study on antileishmanial and antitrypanosomal cinnamate ester analogues. Molecules 24 (23). https://doi.org/10.3390/molecules24234358.
Beverley, S.M., Ellenberger, T.E., Cordingley, J.S., 1986. Primary structure of the gene encoding the bifunctional dihydrofolate reductase-thymidylate synthase of Leishmania major. Proc. Natl. Acad. Sci. U. S. A. 83 (8), 2584–2588. https://doi.org/10.1073/pnas.83.8.2584.
Blanco-González, A., et al., 2023. The role of AI in drug discovery: challenges, opportunities, and strategies. Pharmaceuticals 16 (6), 1–11. https://doi.org/10.3390/ph16060891.
Bora, K., Sarma, M., Kanaujia, S.P., Dubey, V.K., 2023. Dual-target drugs against Leishmania donovani for potential novel therapeutics. Sci. Rep. 13 (1), 1–15. https://doi.org/10.1038/s41598-023-45448-x.
Brannigan, J.A., Wilkinson, A.J., 2021. Drug discovery in leishmaniasis using protein lipidation as a target. Biophys. Rev. 13 (6), 1139–1146. https://doi.org/10.1007/s12551-021-00855-0.
Breslin, W., Pham, D., 2023. Machine learning and drug discovery for neglected tropical diseases. BMC Bioinforma. 24 (1), 165. https://doi.org/10.1186/s12859-022-05076-0.
Bulté, D., et al., 2021. Miltefosine enhances infectivity of a miltefosine-resistant leishmania infantum strain by attenuating its innate immune recognition. PLoS Negl. Trop. Dis. 15 (7), 1–27. https://doi.org/10.1371/journal.pntd.0009622.
Buterez, D., Janet, J.P., Kiddle, S.J., Oglic, D., Lió, P., 2024. Transfer learning with graph neural networks for improved molecular property prediction in the multi-fidelity setting. Nat. Commun. 15 (1), 1–18. https://doi.org/10.1038/s41467-024-45566-8.
Cabitza, F., et al., 2021. The importance of being external. methodological insights for the external validation of machine learning models in medicine. Comput. Methods Prog. Biomed. 208, 106288. https://doi.org/10.1016/j.cmpb.2021.106288.
Cai, C., et al., 2020. Transfer learning for drug discovery. J. Med. Chem. 63 (16), 8683–8694. https://doi.org/10.1021/acs.jmedchem.9b02147.
Capecchi, A., Probst, D., Reymond, J.L., 2020. One molecular fingerprint to rule them all: Drugs, biomolecules, and the metabolome. J. Chemin.-. 12 (1), 1–15. https://doi.org/10.1186/s13321-020-00445-4.
Cereto-Massagué, A., Ojeda, M.J., Valls, C., Mulero, M., Garcia-Vallvé, S., Pujadas, G., 2015. Molecular fingerprint similarity search in virtual screening (no. C). Methods 71, 58–63. https://doi.org/10.1016/j.ymeth.2014.08.005.
Colby, L.A., Quenee, L.E., Zitzow, L.A., 2017. Considerations for infectious disease research studies using animals. Comp. Med. 67 (3), 222–231.
Croft, S.L., Coombs, G.H., 2003. Leishmaniasis - current chemotherapy and recent advances in the search for novel drugs. Trends Parasitol. 19 (11), 502–508. https://doi.org/10.1016/j.pt.2003.09.008.
Croft, S.L., Engel, J., 2006. Miltefosine - discovery of the antileishmanial activity of phospholipid derivatives. Trans. R. Soc. Trop. Med. Hyg. 100 (. 1), 4–8. https://doi.org/10.1016/j.trstmh.2006.03.009.
Cui, H., Cheetham, A.G., Pashuck, E.T., Stupp, S.I., 2014. Amino acid sequence in constitutionally isomeric tetrapeptide amphiphiles dictates architecture of one-dimensional nanostructures. J. Am. Chem. Soc. 136 (35), 12461–12468. https://doi.org/10.1021/ja507051w.
Cunningham, M.L., Fairlamb, A.H., 1995. Trypanothione reductase from Leishmania donovani purification, characterisation and inhibition by trivalent antimonials. Eur. J. Biochem. 230 (2), 460–468. https://doi.org/10.1111/j.1432-1033.1995.tb20583.x.
David, L., Thakkar, A., Mercado, R., Engkvist, O., 2020. Molecular representations in AI-driven drug discovery: a review and practical guide. J. Chemin.-. 12 (1), 1–22. https://doi.org/10.1186/s13321-020-00460-5.
De Muylder, G., Vanhollebeke, B., Caljon, G., Wolfe, A.R., McKerrow, J., Dujardin, J.C., 2016. Naloxonazine, an amastigote-specific compound, affects leishmania parasites through modulation of host-encoded functions. PLoS Negl. Trop. Dis. 10 (12), 1–14. https://doi.org/10.1371/journal.pntd.0005234.
Doyle, M.A., MacRae, J.I., De Souza, D.P., Saunders, E.C., McConville, M.J., Likić, V.A., 2009. LeishCyc: A biochemical pathways database for Leishmania major. BMC Syst. Biol. 3, 1–12. https://doi.org/10.1186/1752-0509-3-57.
Duncan, R., Gannavaram, S., Dey, R., Debrabant, A., Lakhal-Naouar, I., Nakhasi, H.L., 2011. Identification and characterization of genes involved in leishmania pathogenesis: the potential for drug target selection. Mol. Biol. Int. 2011, 1–10. https://doi.org/10.4061/2011/428486.
Feng, X., Ma, Z., Yu, C., Xin, R., 2024. MRNDR: multihead attention-based recommendation network for drug repurposing. J. Chem. Inf. Model. 64 (7), 2654–2669. https://doi.org/10.1021/acs.jcim.3c01726.
Fernández-De Gortari, E., García-Jacas, C.R., Martinez-Mayorga, K., Medina-Franco, J.L., 2017. Database fingerprint (DFP): an approach to represent molecular databases. J. Chemin.-. 9 (1), 1–9. https://doi.org/10.1186/s13321-017-0195-1.
Ferreira, L.G., Dos Santos, R.N., Oliva, G., Andricopulo, A.D., 2015. Molecular docking and structure-based drug design strategies, 20 (7). https://doi.org/10.3390/molecules200713384.
Francesconi, V., Rizzo, M., Schenone, S., Carbone, A., Tonelli, M., 2023. State-of-the-art review on the antiparasitic activity of benzimidazolebased derivatives: facing malaria, leishmaniasis, and trypanosomiasis. Curr. Med. Chem. 31 (15), 1955–1982. https://doi.org/10.2174/0929867331666230915093928.
Gaulton, A., et al., 2012. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40 (D1), 1100–1107. https://doi.org/10.1093/nar/gkr777.
Goodarzi, M., Dejaegher, B., Heyden, Y.Vander, 2012. Feature selection methods in QSAR studies. J. AOAC Int. 95 (3), 636–651. https://doi.org/10.5740/jaoacint.SGE_Goodarzi.
Guha, R., Willighagen, E., 2020. A survey of quantitative descriptions of molecular structure. Curr. Top. Med. Chem. 12 (18), 1946–1956. https://doi.org/10.2174/1568026611212180002.
Hasebe, T., 2021. Knowledge-embedded message-passing neural networks: improving molecular property prediction with human knowledge. ACS Omega 6 (42), 27955–27967. https://doi.org/10.1021/acsomega.1c03839.
Hu, S., Jiang, S., Qi, X., Bai, R., Ye, X.Y., Xie, T., 2022. Races of small molecule clinical trials for the treatment of COVID-19: an up-to-date comprehensive review. Drug Dev. Res. 83 (1), 16–54. https://doi.org/10.1002/ddr.21895.
Huang, H.H., Wu, N.Q., Liang, Y., Peng, X.D., Shu, J., 2022. SLNL: a novel method for gene selection and phenotype classification. Int. J. Intell. Syst. 37 (9), 6283–6304. https://doi.org/10.1002/int.22844.
Jebbari, H., Davidson, R., 1998. Recent advances in leishmaniasis. Curr. Opin. Infect. Dis. 11 (5), 535–539. https://doi.org/10.1097/00001432-199810000-00004.
Jha, T.K., Sundar, S., Thakur, C.P., Felton, J.M., Sabin, A.J., Horton, J., 2005. A phase II dose-ranging study of sitamaquine for the treatment of visceral leishmaniasis in India. Am. J. Trop. Med. Hyg. 73 (6), 1005–1011. https://doi.org/10.4269/ajtmh.2005.73.1005.
Jiang, D., et al., 2021. Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. J. Chemin.-. 13 (1), 1–23. https://doi.org/10.1186/s13321-020-00479-8.
P.B. Jørgensen, K.W. Jacobsen, and M.N. Schmidt, Neural Message Passing with Edge Updates for Predicting Properties of Molecules and Materials, no. 2005, 2018, [Online]. Available: ⟨http://arxiv.org/abs/1806.03146⟩.
Kim, S., et al., 2016. PubChem substance and compound databases. Nucleic Acids Res 44 (D1), D1202–D1213. https://doi.org/10.1093/nar/gkv951.
Kim, S., Han, L., Yu, B., Hähnke, V.D., Bolton, E.E., Bryant, S.H., 2015. PubChem structure-activity relationship (SAR) clusters. J. Chemin.-. 7 (1), 1–22. https://doi.org/10.1186/s13321-015-0070-x.
Knight, C.A., Harris, D.R., Alshammari, S.O., Gugssa, A., Young, T., Lee, C.M., 2023. Leishmaniasis: Recent epidemiological studies in the Middle East (no. February). Front. Microbiol. 13, 1–13. https://doi.org/10.3389/fmicb.2022.1052478.
Y. Lau, J.M. Gutierrez, M. Volkovs, and S. Zuberi, Drug repurposing for Leishmaniasis with Hyperbolic Graph Neural Networks, *bioRxiv*, p. 2023.02.11.528117, 2023, [Online]. Available: ⟨http://biorxiv.org/content/early/2023/02/13/2023.02.11.528117.abstract⟩.
Li, X., et al., 2024. Screening for primary aldosteronism on and off interfering medications. Endocrine 83 (1), 178–187. https://doi.org/10.1007/s12020-023-03520-6.
Lira, R., et al., 1999. Evidence that the high incidence of treatment failures in Indian kala- azar is due to the emergence of antimony-resistant strains of Leishmania donovani. J. Infect. Dis. 180 (2), 564–567. https://doi.org/10.1086/314896.
T. Liu *et al.*, BindingDBBindingDB in 2024: a FAIR Knowledgebase of Protein-Small Molecule Binding Data, pp. 1–12, 2024, [Online]. Available: ⟨https://chemrxiv.org/engage/chemrxiv/article-details/66e360db12ff75c3a12b3a00⟩.
Maeda, M.H., 2015. Current challenges in development of a database of three-dimensional chemical structures. Front. Bioeng. Biotechnol. 3 (MAY), 1–9. https://doi.org/10.3389/fbioe.2015.00066.
Majumder, N., Banerjee, A., Saha, S., 2023. A review on new natural and synthetic anti-leishmanial chemotherapeutic agents and current perspective of treatment approaches. no. November 2022, p Acta Trop. 240, 106846. https://doi.org/10.1016/j.actatropica.2023.106846.
Mazón-Ortiz, G., Cerda-Mejía, G., Gutiérrez Morales, E., Diéguez-Santana, K., Ruso, J.M., González-Díaz, H., 2023. Trends in nanoparticles for leishmania treatment: a bibliometric and network analysis. Diseases 11 (4). https://doi.org/10.3390/diseases11040153.
McGibbon, M., et al., 2024. From intuition to AI: evolution of small molecule representations in drug discovery. Brief. Bioinform. 25 (1), 1–13. https://doi.org/10.1093/bib/bbad422.
Mercado-Camargo, J., et al., 2020. Homology Modeling of Leishmanolysin (gp63) from leishmania panamensis and molecular docking of flavonoids. ACS Omega 5 (24), 14741–14749. https://doi.org/10.1021/acsomega.0c01584.

Mirza, H., Teo, J.D.W., Upcroft, J., Tan, K.S.W., 2011. A rapid, high-throughput viability assay for Blastocystis spp. reveals metronidazole resistance and extensive subtype-dependent variations in drug susceptibilities. Antimicrob. Agents Chemother. 55 (2), 637–648. https://doi.org/10.1128/AAC.00900-10.

Mollalo, A., Sadeghian, A., Israel, G.D., Rashidi, P., Sofizadeh, A., Glass, G.E., 2018. Machine learning approaches in GIS-based ecological modeling of the sand fly Phlebotomus papatasi, a vector of zoonotic cutaneous leishmaniasis in Golestan province, Iran (no. July). Acta Trop. 188, 187–194. https://doi.org/10.1016/j.actatropica.2018.09.004.

Moon, S., Lee, H., 2021. JDSNMF: Joint deep semi-non-negative matrix factorization for learning integrative representation of molecular signals in Alzheimer's disease. J. Pers. Med. 11 (8), 1–16. https://doi.org/10.3390/jpm11080686.

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

Nguyen, L.C., Nguyen-Xuan, H., 2020. Deep learning for computational structural optimization (no. xxxx). ISA Trans. 103, 177–191. https://doi.org/10.1016/j.isatra.2020.03.033.

O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R., 2011. Open Babel. J. Chemin.-. 3 (33), 1–14. ⟨https://jcheminf.biomedcentral.com/track/pdf/10.1186/1758-2946-3-33⟩ (Online]. Available).

Ochoa, R., Watowich, S.J., Flórez, A., Mesa, C.V., Robledo, S.M., Muskus, C., 2016. Drug search for leishmaniasis: a virtual screening approach by grid computing. J. Comput. Aided Mol. Des. 30 (7), 541–552. https://doi.org/10.1007/s10822-016-9921-4.

Okwor, I., Uzonna, J., 2016. Social and economic burden of human leishmaniasis. Am. J. Trop. Med. Hyg. 94 (3), 489–493. https://doi.org/10.4269/ajtmh.15-0408.

Oualha, R., Abdelkrim, Y.Z., Guizani, I., Harigua-Souiai, E., 2024. Approved drugs successfully repurposed against Leishmania based on machine learning predictions (no. September, p). Front. Cell. Infect. Microbiol. 14, 1403589. https://doi.org/10.3389/fcimb.2024.1403589.

Pérez-Victoria, F.J., et al., 2006. Mechanisms of experimental resistance of Leishmania to miltefosine: Implications for clinical use. Drug Resist. Updat. 9 (1–2), 26–39. https://doi.org/10.1016/j.drup.2006.04.001.

Ponzoni, I., et al., 2017. Hybridizing Feature Selection and Feature Learning Approaches in QSAR Modeling for Drug Discovery /631/114/2248 /631/154/309 /639/638/563/606 /119/118 article. Sci. Rep. 7 (1), 1–19. https://doi.org/10.1038/s41598-017-02114-3.

Prosser, K.E., Stokes, R.W., Cohen, S.M., 2020. Evaluation of 3-dimensionality in approved and experimental drug space. ACS Med. Chem. Lett. 11 (6), 1292–1298. https://doi.org/10.1021/acsmedchemlett.0c00121.

Purkait, B., et al., 2012. Mechanism of amphotericin B resistance in clinical isolates of Leishmania donovani. Antimicrob. Agents Chemother. 56 (2), 1031–1041. https://doi.org/10.1128/AAC.00030-11.

Rabaan, A.A., et al., 2023. Omics approaches in drug development against leishmaniasis: current scenario and future prospects. Pathogens 12 (1), 1–20. https://doi.org/10.3390/pathogens12010039.

Rahi, A.A., 2013. Comparison of diagnostic methods in cutaneous leishmaniasis in Iraq. Am. J. Biosci. 1 (1), 1. https://doi.org/10.11648/j.ajbio.20130101.11.

Ramezankhani, R., Sajjadi, N., Esmaeilzadeh, R.N., Jozi, S.A., Shirzadi, M.R., 2018. Application of decision tree for prediction of cutaneous leishmaniasis incidence based on environmental and topographic factors in Isfahan province, Iran. Geospat. Health 13 (1), 172–178. https://doi.org/10.4081/gh.2018.664.

Reimão, J.Q., Scotti, M.T., Tempone, A.G., 2010. Anti-leishmanial and anti-trypanosomal activities of 1,4-dihydropyridines: In vitro evaluation and structure-activity relationship study. Bioorg. Med. Chem. 18 (22), 8044–8053. https://doi.org/10.1016/j.bmc.2010.09.015.

Rezaei, Z., Tahmasebi, A., Pourabbas, B., 2024. Using meta-analysis and machine learning to investigate the transcriptional response of immune cells to Leishmania infection. PLoS Negl. Trop. Dis. 18 (1), 1–18. https://doi.org/10.1371/journal.pntd.0011892.

Roatt, B.M., et al., 2014. Immunotherapy and immunochemotherapy in visceral leishmaniasis: Promising treatments for this neglected disease. Front. Immunol. 5, 1–12. https://doi.org/10.3389/fimmu.2014.00272.

Robles-Loaiza, A.A., et al., 2021. Peptides to tackle leishmaniasis: current status and future directions. Int. J. Mol. Sci. 22 (9). https://doi.org/10.3390/ijms22094400.

Rohini, K., Ramanathan, K., Shanthi, V., 2019. Multi-dimensional screening strategy for drug repurposing with statistical framework—a new road to Influenza Drug discovery. Cell Biochem. Biophys. 77 (4), 319–333. https://doi.org/10.1007/s12013-019-00887-0.

Ruggeri, B.A., Camp, F., Miknyoczki, S., 2014. Animal models of disease: Pre-clinical animal models of cancer and their applications and utility in drug discovery. Biochem. Pharmacol. 87 (1), 150–161. https://doi.org/10.1016/j.bcp.2013.06.020.

Saini, I., Joshi, J., Kaur, S., 2022. Unwelcome prevalence of leishmaniasis with several other infectious diseases. Int. Immunopharmacol. 110, 109059. https://doi.org/10.1016/j.intimp.2022.109059.

Sasidharan, S., Saudagar, P., 2020. Mapping N- and C-terminals of Leishmania donovani tyrosine aminotransferase by gene truncation strategy: a functional study using in vitro and in silico approaches. Sci. Rep. 10 (1), 1–15. https://doi.org/10.1038/s41598-020-69512-y.

Schulz-Gasch, T., Schärfer, C., Guba, W., Rarey, M., 2012. TFD: Torsion fingerprints as a new measure to compare small molecule conformations. J. Chem. Inf. Model. 52 (6), 1499–1512. https://doi.org/10.1021/ci2002318.

Sebastián-Pérez, V., et al., 2022. Naphthoquinone as a New Chemical Scaffold for Leishmanicidal Inhibitors of Leishmania GSK-3. Biomedicines 10 (5). https://doi.org/10.3390/biomedicines10051136.

Seifert, K., 2011. Structures, targets and recent approaches in anti-leishmanial drug discovery and development. Open Med. Chem. J. 5, 31–39. https://doi.org/10.2174/1874104501105010031.

Shah, H.A., Liu, J., Yang, Z., Zhang, X., Feng, J., 2022. DeepRF: a deep learning method for predicting metabolic pathways in organisms based on annotated genomes (no. June, p). Comput. Biol. Med. 147, 105756. https://doi.org/10.1016/j.compbiomed.2022.105756.

Shah, H.A., Liu, J., Yang, Z., Yang, F., Zhang, Q., Feng, J., 2023. DeepRT: predicting compounds presence in pathway modules and classifying into module classes using deep neural networks based on molecular properties. J. Bioinform. Comput. Biol. 21 (4), 1–20. https://doi.org/10.1142/S0219720023500178.

Shah, H.A., Liu, J., Yang, Z., 2024. Gtie-Rt: a comprehensive graph learning model for predicting drugs targeting metabolic pathways in human. J. Bioinform. Comput. Biol. 22 (3), 1–21. https://doi.org/10.1142/S0219720024500100.

Shi, D., Mi, G., Wang, M., Webster, T.J., 2019. In vitro and ex vivo systems at the forefront of infection modeling and drug discovery. no. October 2018 Biomaterials 198, 228–249. https://doi.org/10.1016/j.biomaterials.2018.10.030.

S. Sifontes-rodríguez, S. Meneses-gómez, A.R. Escalona-montaño, D.A. Sánchez-almaraz, and O. Pérez-olvera, PubChem BioAssays 1063: A poorly exploited source of new antileishmanial compounds, pp. 1–21.

Singh, P., Kumar, A., 2020. Deciphering the function of unknown Leishmania donovani cytosolic proteins using hyperparameter-tuned random forest. Netw. Model. Anal. Heal. Inform. Bioinforma. 9 (1), 1–9. https://doi.org/10.1007/s13721-019-0208-2.

Soiza, R.L., Donaldson, A.I.C., Myint, P.K., 2018. Vaccine against arteriosclerosis: an update. Ther. Adv. Vaccin. 9 (6), 259–261. https://doi.org/10.1177/https.

Soni, M., Pratap, J.V., 2022. Development of novel anti-leishmanials: the case for structure-based approaches. Pathogens 11 (8). https://doi.org/10.3390/pathogens11080950.

Sundar, S., Chatterjee, M., 2006. Visceral leishmaniasis - current therapeutic modalities. Indian J. Med. Res. 123 (3), 345–352.

K. Swanson, Message Passing Neural Networks for Molecular Property Prediction by, 2019, [Online]. Available: https://hdl.handle.net/1721.1/123133.

Tegazzini, D., et al., 2016. A replicative in vitro assay for drug discovery against Leishmania donovani. Antimicrob. Agents Chemother. 60 (6), 3524–3532. https://doi.org/10.1128/AAC.01781-15.

Turcano, L., et al., 2018. Identification and binding mode of a novel Leishmania Trypanothione reductase inhibitor from high throughput screening. PLoS Negl. Trop. Dis. 12 (11), 1–21. https://doi.org/10.1371/journal.pntd.0006969.

Ugbe, F.A., Shallangwa, G.A., Uzairu, A., Abdulkadir, I., 2022. Theoretical activity prediction, structure-based design, molecular docking and pharmacokinetic studies of some maleimides against Leishmania donovani for the treatment of leishmaniasis. Bull. Natl. Res. Cent. 46 (1). https://doi.org/10.1186/s42269-022-00779-z.

Upegui Zapata, Y.A., et al., 2020. Mode of action of a formulation containing hydrazones and saponins against leishmania spp. Role in mitochondria, proteases and reinfection process. Int. J. Parasitol. Drugs Drug Resist. 13, 94–106. https://doi.org/10.1016/j.ijpddr.2020.06.004.

Vlachakis, D., Vlamos, P., 2021. Mathematical multidimensional modelling and structural artificial intelligence pipelines provide insights for the designing of highly specific AntiSARS-CoV2 agents. Math. Comput. Sci. 15 (4), 877–888. https://doi.org/10.1007/s11786-021-00517-0.

Wang, X., Chen, G., Hu, H., Zhang, M., Rao, Y., Yue, Z., 2024. PDDGCN: a parasitic disease–drug association predictor based on multi-view fusion graph convolutional network. Interdiscip. Sci. – Comput. Life Sci. 16 (1), 231–242. https://doi.org/10.1007/s12539-023-00600-z.

Winkler, D.A., 2021. Use of artificial intelligence and machine learning for discovery of drugs for neglected tropical diseases. Front. Chem. 9, 1–15. https://doi.org/10.3389/fchem.2021.614073.

Wishart, D.S., et al., 2008. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 36 (. 1), 901–906. https://doi.org/10.1093/nar/gkm958.

Yang, Z., Liu, J., Shah, H.A., Feng, J., 2022. A novel hybrid framework for metabolic pathways prediction based on the graph attention network. BMC Bioinforma. 23, 1–14. https://doi.org/10.1186/s12859-022-04856-y.

Yardley, V., Croft, S.L., 2000. A comparison of the activities of three amphotericin B lipid formulations against experimental visceral and cutaneous leishmaniasis. Int. J. Antimicrob. Agents 13 (4), 243–248. https://doi.org/10.1016/S0924-8579(99)00133-8.

Ye, M., Zhou, Y., Zhao, H., Wang, X., 2023. Magnetic microrobots with folate targeting for drug delivery. Cyborg Bionic Syst. 4. https://doi.org/10.34133/cbsystems.0019.

Yin, D., et al., 2024. Dendritic-cell-targeting virus-like particles as potent mRNA vaccine carriers. Nat. Biomed. Eng. https://doi.org/10.1038/s41551-024-01208-4.

Zare, M., et al., 2022. A machine learning-based system for detecting leishmaniasis in microscopic images. BMC Infect. Dis. 22 (1), 1–6. https://doi.org/10.1186/s12879-022-07029-7.

Zdrazil, B., et al., 2024. The ChEMBL Database in 2023:ā drug disco v ery platf orm spanning multiple bioactivity data typesānd time periods. Nucleic Acids Res 52 (D1), D1180–D1192. https://doi.org/10.1093/nar/gkad1004.

Zhang, Q.Y., et al., 2024. Multi targeted therapy for alzheimer's disease by guanidinium-modified calixarene and cyclodextrin co-assembly loaded with insulin. ACS Nano. https://doi.org/10.1021/acsnano.4c05693.

Zhao, B.W., et al., 2023. Fusing higher and lower-order biological information for drug repositioning via graph representation learning. IEEE Trans. Emerg. Top. Comput. 14 (8). https://doi.org/10.1109/TETC.2023.3239949.

Zheng, P., et al., 2024. A multi-omic analysis reveals that Gamabufotalin exerts anti-hepatocellular carcinoma effects by regulating amino acid metabolism through targeting STAMBPL1. no. 283, p Phytomedicine 135, 156094. https://doi.org/10.1016/j.phymed.2024.156094.

Zhu, J., et al., 2023. Combination of chemotherapy and gaseous signaling molecular therapy: Novel β-elemene nitric oxide donor derivatives against leukemia. Drug Dev. Res. 84 (4), 718–735. https://doi.org/10.1002/ddr.22051.