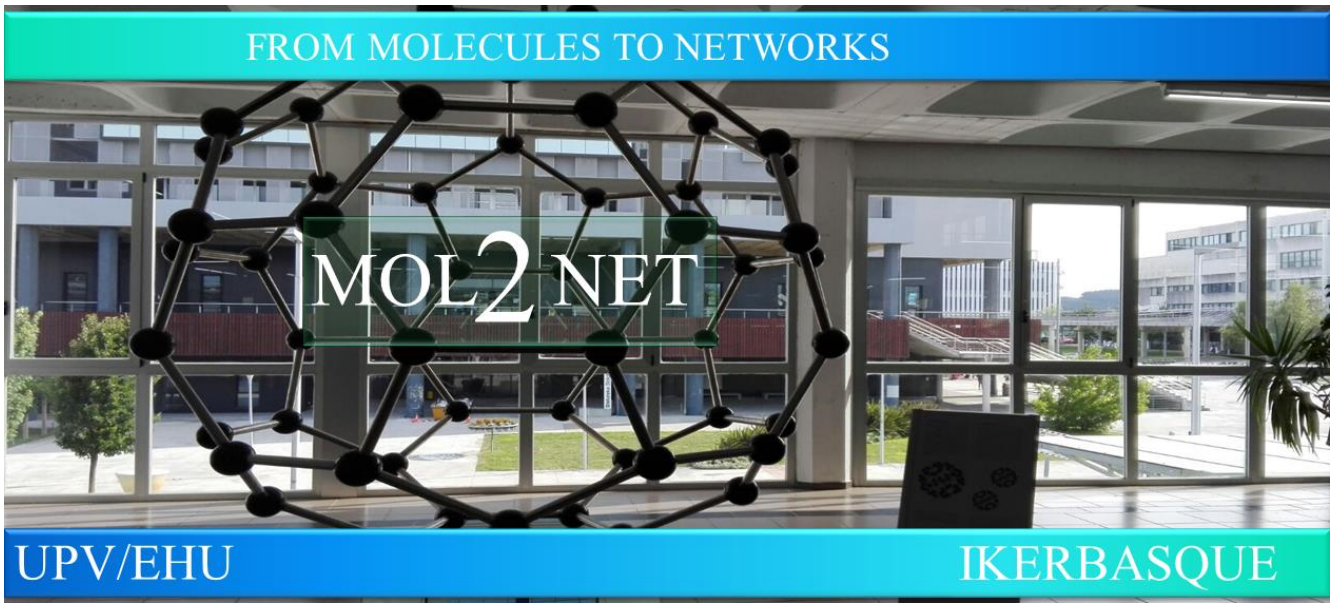
MDPI sciforum

MOL2NET'22, Conference on Molecular, Biomedical & Computational Sciences and Engineering, 8th ed.



Artificial intelligence and machine learning in Leishmania drug discovery

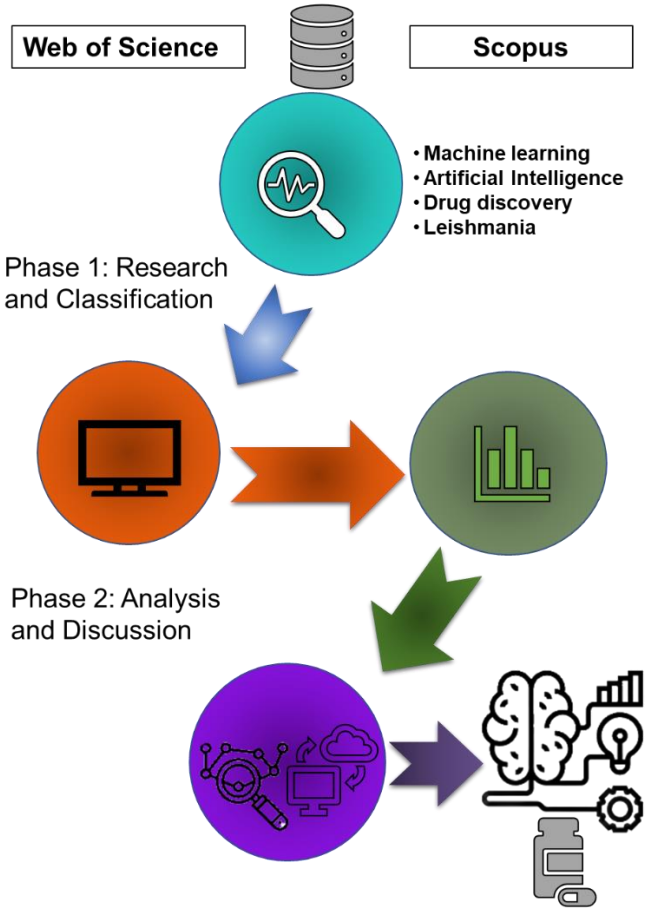
Gabriel Mazón-Ortiz 1 [a,b], Galo Cerda-Mejía 2 [a,b], Karel Diéguez-Santana 3 [a], Juan M. Ruso 4 [b]

[a] Universidad Regional Amazónica Ikiam, Parroquia Muyuna km 7 vía Alto Tena, 150150, Tena-Napo, Ecuador. E-mail: karel.dieguez.santana@gmail.com

[b] Soft Matter and Molecular Biophysics Group, Department of Applied Physics, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain

.

Graphical Abstract	**Abstract.**
	Leishmaniasis is a vector-borne parasite that affects 700,000–1 million people and kills 26,000–65,000 annually. This work presents a brief, comprehensive review of AI/ML studies



performed in leishmania drug discovery. In this study, research was carried out using the Scopus and Web of Science (WoS) databases from 2013 to 2022. Between the two databases, 28 documents were found, eight of which were duplicates; hence, a total of 20 articles were analysed. Of these, nine were research articles, 10 review articles, and one editorial document. Current Topics in Medicinal Chemistry was the only journal that received more than one paper (three papers in total). The available literature on the topic is limited. The most relevant articles selected from the two databases, WoS and Scopus, provided an overview of the scientific topic (stages of drug discovery for leishmaniasis using AI/ML). This brief review can also serve as a starting point for integrating knowledge in this field through research and suggest future research avenues for AI and ML applications in protozoan infectious diseases.

Introduction

Leishmaniasis is caused by ~20 Leishmania species and transmitted by promastigote-infected female phlebotomine sandflies during a blood meal. In the mammalian host, promastigotes are absorbed by phagocytic cells, convert into amastigotes, rapidly multiply, and infect the vector through the next blood meal [1]. The three primary diseases have different symptoms and severity levels. Visceral leishmaniasis (VL)—known as kala-azar in India—is the deadliest type of illness [2]. The most common form of leishmaniasis, cutaneous leishmaniasis (CL), causes skin lesions and ulcers that can leave lifelong scars and severe disabilities; mucocutaneous leishmaniasis destroys tissues in the nose, mouth, and throat [3]. Leishmaniasis symptoms range from mild in CL to severe in VL. Algeria, Afghanistan, Bolivia, Brazil, China, Colombia, Eritrea, Ethiopia, India, Iraq, Kenya, Libya, Pakistan, Somalia, South Sudan, Sudan, the Syrian Arab Republic, Tunisia, and Yemen account for ~90% of infections.

Leishmaniasis is a vector-borne parasite that affects 700,000–1 million people and kills 26,000–65,000 annually. These disorders lack low-cost, non-toxic therapies, prompting novel therapeutic options. High attrition, rising expenses, and pricing pressure plague drug R&D. Neglected tropical diseases (NTDs)—including leishmaniases—are the least prioritised diseases for therapeutic development due to their low market potential [4].

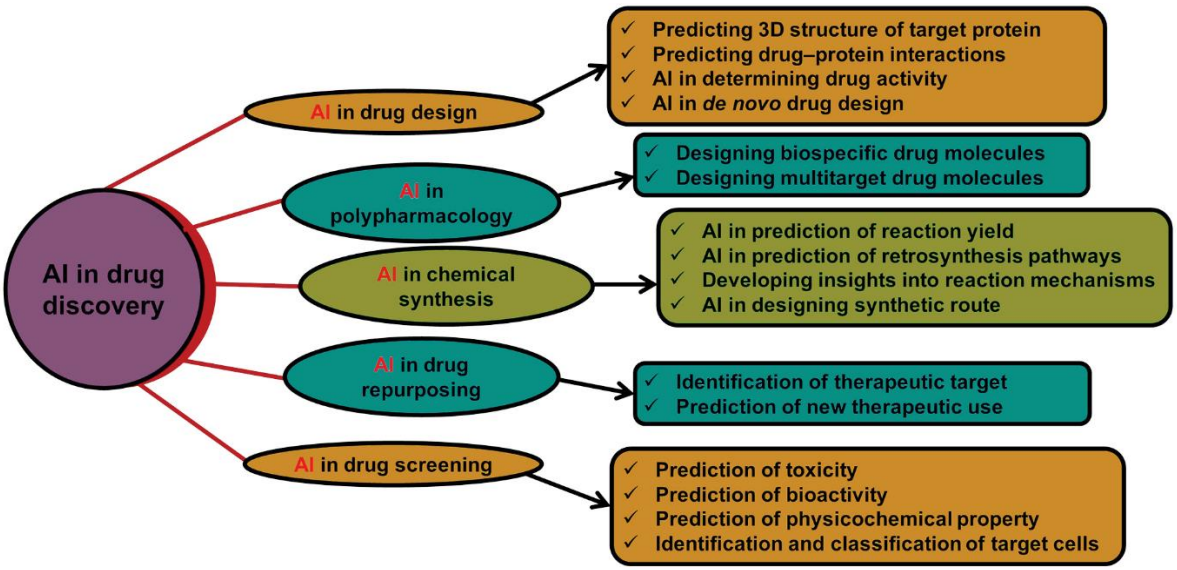
Leishmaniasis has been found in Mexico, Central America, South America, and Texas and Oklahoma in the US. Leishmaniasis epidemiology depends on parasite and sandfly species, infection site ecology, human physiology, and parasitic infection history [3]. Most leishmaniasis chemotherapeutic treatments have severe drawbacks [5]. Most anti-leishmanial treatments require multiple doses over a lengthy period, challenging clinical leishmaniasis management and promoting drug resistance [6,7].

On the other hand, academic and private researchers find drug discovery (DD) and lead optimisation difficult. As the number of approved pharmaceuticals decreases, the development costs of a novel, viable medicinal molecule approach US $1.2 billion for a 12-year research process. Computational methods could reduce DD cost, duration, and attrition [8,9]. Several groups have developed drug research and algorithms for artificial intelligence (AI) and machine learning (ML) [10,11]. DD and lead optimisation need more data, and few compounds reach the market. Due to insufficient data, AI/ML methods were developed, including one-shot learning based on structure–activity connections for activity predictions [12].

In the case of NTDs, we have even fewer data. These diseases mainly afflict the poor and vulnerable, and extracellular tests and primary high-throughput screenings are the main sources of drug discovery data for such disorders. It may not access vast datasets of compounds validated at various experimental levels (enzymatic, *in vitro*, *in vivo*, *ex vivo*, *etc*.). Early-stage drug development data has to be used to develop cost-effective medication discovery and repurposing methods for such disorders [4].

In this sense, AI/ML have been employed in various stages of drug discovery, for example, de novo drug design, chemical synthesis, virtual screening, and in silico evaluation of the properties (absorption, distribution, metabolism, excretion, and toxicity) of a drug molecule. Figure 1 shows the main applications of AI/ML in drug discovery. AI can recognise hit and lead compounds and provide faster drug target validation and drug structure design optimisation [13]. QSAR modelling tools have been used for the identification of potential drug candidates [14]. They have evolved into AI-based QSAR approaches, such as linear discriminant analysis (LDA), support vector machines (SVM), neural networks (NN), random forests (RF), and decision trees, which can be applied to accelerate QSAR analysis [15-18].

Figure 1. The Role of AI/ML in Drug Discovery.



AI/ML can be used successfully in drug discovery, including drug design, polypharmacology, chemical synthesis, drug repurposing, and drug screening (reproduced from [19,20]).

In this work, we present a brief comprehensive review of AI/ML studies performed in leishmania drug discovery.

Materials and Methods

In this study, research was carried out using the Scopus and Web of Science (WoS) databases, which are the two main international academic information sources. Scopus data were manually extracted from the online version (www.scopus.com) and WoS data from the Web of Science Core Collection publications. For the publication time span, the time from 2013 to date (December 6, 2022) was considered with the intention of understanding how research into the use of AI/ML for drug discovery to treat leishmania has evolved. The research methodology chosen for this study was a systematic review of the literature, and the main phases of the study were as follows:

Phase 1: Research and Classification. This phase was divided into three steps: identification, screening, and inclusion. Documents were searched, a screening of the overall output was carried out to identify which documents could be taken into account, in line with the research areas considered interesting and relevant, and documents were selected to be analysed in detail.

Phase 2: Analysis and Discussion. In this phase the analysis of the results was carried out. The results were then discussed and conclusions were drawn.

.

Results and Discussion

. The search was conducted using various terms related to AI/ML, Leishmania, drug discovery, and with Boolean operators, and the following results were obtained:

Scopus analysis

Search strategy: (TITLE-ABS-KEY (artificial AND intelligence) OR TITLE-ABS-KEY (machine AND learning) AND TITLE-ABS-KEY (leishmania) AND TITLE-ABS-KEY (drug AND discovery))

Results: 13 document results.

WoS Analysis (Web of Science Core Collection)

Search strategy: TS = (machine learning, leishmania, drug discovery)

Results: 15 document results.

In the analysis of the results of the two databases, eight documents were found to be duplicates, leaving 20 documents in total to be analysed.

Of these, nine research articles, ten reviews, and one editorial from 18 sources, the journal Current Topics in Medicinal Chemistry was the only one that submitted more than one document (three documents).

Finally, all the documents found, after excluding the repeated ones, were as follows: [4,21-39]. After reading the abstracts, the articles presented in **Table 1** were selected. Twelve articles were excluded, one of them because it was an editorial document, and the rest because they were focused on general studies and related to other parasitic diseases.

Table 1. Published studies are AI/ML in drug discovery against Leishmania

No	Title	Year	Description	Journal	Ref.
1	Development of Novel Anti-Leishmanials: The Case for	2022	Summarises the disease epidemiology and available therapies, we consider three important leishmanial metabolic pathways that can be attractive targets	Pathogens	[31]

	Structure-Based Approaches.		for a structure-based drug discovery approach towards the development of novel anti-leishmanials.		
2	Machine Learning and Its Applications for Protozoal Pathogens and Protozoal Infectious Diseases.	2022	Brief overview of important concepts in ML serving as background knowledge (infectious diseases caused by protozoal pathogens), with a focus on basic workflows, popular algorithms (*e.g.*, SVM, RF, and NN), feature extraction and selection, and model evaluation metrics.	Frontiers in Cellular and Infection Microbiology	[26]
3	Applied Machine Learning Toward Drug Discovery Enhancement: Leishmaniases as a Case Study.	2022	They used four machine learning (ML) algorithms that were trained and tested for their ability to classify molecules into active or inactive classes based on the chemical structure of various molecular fingerprints (FPs) of 65 057 molecules against *Leishmania major* promastigotes.	Bioinformatics and Biology Insights	[4]
4	Palladium-mediated synthesis and biological evaluation of C-10b substituted Dihydropyrrolo[1,2-b]isoquinolines as antileishmanial agents.	2021	A series of synthesised molecules (C-10 substituted derivatives) were tested in vitro for leishmanicidal activity against visceral (*L. donovani*) and cutaneous (*L. amazonensis*) leishmaniasis. In addition, they developed the first perturbation theory machine learning algorithm capable of simultaneously predicting multiple parameters of biological activity against any Leishmania species and target protein.	European Journal of Medicinal Chemistry	[23]
5	Discovery of Alternative Chemotherapy Options for Leishmaniasis through Computational Studies of Asteraceae.	2021	Reviews various studies that have used computational tools to examine various compounds identified in the Asteraceae family in the search for potential drug candidates against Leishmania.	ChemMedChem	[24]
6	Colombian contributions fighting leishmaniasis: A systematic review on antileishmanials combined with	2020	They reviewed Colombian antileishmanial research studies in order to combine general efforts aimed at finding a lead against *Leishmania panamensis* and to recognise the structural characteristics of	Molecules	[30]

	chemoinformatics analysis.		representative compounds. They found that fingerprint-based analyses are available using conventional machine learning algorithms and clustering methods.		
7	Computational Identification of Chemical Compounds with Potential Activity against *Leishmania amazonensis* using Nonlinear Machine Learning Techniques.	2018	They develop computational models for the identification of new chemical compounds with potential antileishmanial activity. They used a set of 116 organic chemical compounds that had been tested against promastigotes of *Leishmania amazonensis* to make the theoretical models.	Current Topics in Medicinal Chemistry	[33]
8	Cheminformatic models based on machine learning for pyruvate kinase inhibitors of *Leishmania Mexicana.*	2013	They use the machine learning approach to create computational models capable of predicting the biological activity of novel antileishmanial compounds. They analyse a publicly available high-throughput screening dataset of chemical molecules that have been considered targets of the *L. mexicana* pyruvate kinase (LmPK) enzyme. Molecules are evaluated using the substructure-based approach to identify common substructures that contribute to their activity.	BMC Bioinformatics	[27]

.Conclusions

This research focused on a brief survey of the state of the art of AI and ML applications for drug discovery for leishmaniasis. The available literature on the subject is still limited. The most relevant papers were selected from two databases, WoS and Scopus, and an overview of the scientific subject was provided. Furthermore, a starting point for integrating knowledge through research in this area and suggesting future research avenues for AI/ML applications in the field of protozoan infectious diseases was proposed. Finally, it is important to underline that this paper was developed using two databases, but that there are other indexing databases, such as Google Scholar, that could be integrated for future research.

References

1. van Griensven, J.; Ritmeijer, K.; Lynen, L.; Diro, E. Visceral leishmaniasis as an AIDS defining condition: towards consistency across WHO guidelines PLoS neglected tropical diseases**,** 8 (2014**)** e2916, doi:10.1371/journal.pntd.0002916.
2. Zijlstra, E.E. Biomarkers in Post-kala-azar Dermal Leishmaniasis Front Cell Infect Microbiol**,** 9 (2019**)** 228, doi:10.3389/fcimb.2019.00228.

3. WHO. *Weekly Epidemiological Record, 2022, vol. 97, 07 [full issue]*; World Health Organization = Organisation mondiale de la Santé: 2022.
4. Harigua-Souiai, E.; Oualha, R.; Souiai, O.; Abdeljaoued-Tej, I.; Guizani, I. Applied Machine Learning Toward Drug Discovery Enhancement: Leishmaniases as a Case Study Bioinformatics and Biology Insights**,** 16 (2022**)**, doi:10.1177/11779322221090349.
5. Kedzierski, L.; Sakthianandeswaren, A.; Curtis, J.M.; Andrews, P.C.; Junk, P.C.; Kedzierska, K. Leishmaniasis: current treatment and prospects for new drugs and vaccines Current medicinal chemistry**,** 16 (2009**)** 599-614, doi:10.2174/092986709787458489.
6. Andrade Neto, V.V.; Cunha Junior, E.F.; Faioes, V.d.S.; Martins, T.P.; Silva, R.L.; Leon, L.L.; Santos, E.C.T. Leishmaniasis treatment: update of possibilities for drug repurposing Front. Biosci**,** 23 (2018**)** 967–996, doi:10.2741/4629.
7. Roatt, B.M.; de Oliveira Cardoso, J.M.; De Brito, R.C.F.; Coura-Vital, W.; de Oliveira Aguiar-Soares, R.D.; Reis, A.B. Recent advances and new strategies on leishmaniasis treatment Applied Microbiology and Biotechnology**,** 104 (2020**)** 8965-8977, doi:10.1007/s00253-020-10856-w.
8. Diéguez-Santana, K.; González-Díaz, H. Towards machine learning discovery of dual antibacterial drug–nanoparticle systems Nanoscale**,** 13 (2021**)** 17854-17870, doi:10.1039/D1NR04178A.
9. Diéguez-Santana, K.; Rasulev, B.; González-Díaz, H. Towards rational nanomaterial design by predicting drug–nanoparticle system interaction vs. bacterial metabolic networks Environmental Science: Nano**,** 9 (2022**)** 1391-1413, doi:10.1039/D1EN00967B.
10. Varnek, A.; Baskin, I. Machine learning methods for property prediction in chemoinformatics: Quo Vadis? Journal of chemical information and modeling**,** 52 (2012**)** 1413-1437, doi:10.1021/ci200409x.
11. Lo, Y.C.; Rensi, S.E.; Torng, W.; Altman, R.B. Machine learning in chemoinformatics and drug discovery Drug discovery today**,** 23 (2018**)** 1538-1546, doi:10.1016/j.drudis.2018.05.010.
12. Altae-Tran, H.; Ramsundar, B.; Pappu, A.S.; Pande, V. Low Data Drug Discovery with One-Shot Learning ACS central science**,** 3 (2017**)** 283-293, doi:10.1021/acscentsci.6b00367.
13. Mak, K.-K.; Pichika, M.R. Artificial intelligence in drug development: present status and future prospects Drug discovery today**,** 24 (2019**)** 773-780, doi:10.1016/j.drudis.2018.11.014.
14. Dieguez-Santana, K.; M Rivera-Borroto, O.; Puris, A.; Le-Thi-Thu, H.; M Casanola-Martin, G. A Two QSAR Way for Antidiabetic Agents Targeting Using α-Amylase and α-Glucosidase Inhibitors: Model Parameters Settings in Artificial Intelligence Techniques Letters in Drug Design & Discovery**,** 14 (2017**)** 862-868, doi:10.2174/1570180814666161128121142.
15. Wang, Y.; Guo, Y.; Kuang, Q.; Pu, X.; Ji, Y.; Zhang, Z.; Li, M. A comparative study of family-specific protein-ligand complex affinity prediction based on random forest approach Journal of computer-aided molecular design**,** 29 (2015**)** 349-360, doi:10.1007/s10822-014-9827-y.
16. Diéguez-Santana, K.; Casañola-Martin, G.M.; Torres, R.; Rasulev, B.; Green, J.R.; González-Díaz, H. Machine Learning Study of Metabolic Networks vs ChEMBL Data of Antibacterial Compounds Molecular Pharmaceutics**,** 19 (2022**)** 2151-2163, doi:10.1021/acs.molpharmaceut.2c00029.
17. Diéguez-Santana, K.; Rivera-Borroto, O.M.; Puris, A.; Pham-The, H.; Le-Thi-Thu, H.; Rasulev, B.; Casañola-Martin, G.M. Beyond model interpretability using LDA and decision trees for α-amylase and α-glucosidase inhibitor classification studies Chemical biology drug design**,** 94 (2019**)** 1414-1421, doi:10.1111/cbdd.13518.
18. Pham-The, H.; Casañola-Martin, G.; Diéguez-Santana, K.; Nguyen-Hai, N.; Ngoc, N.T.; Vu-Duc, L.; Le-Thi-Thu, H. Quantitative structure–activity relationship analysis and virtual screening studies for identifying HDAC2 inhibitors from known HDAC bioactive chemical libraries SAR and QSAR in Environmental Research**,** 28 (2017**)** 199-220, doi:10.1080/1062936X.2017.1294198.
19. Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R.K. Artificial intelligence in drug discovery and development Drug discovery today**,** 26 (2021**)** 80-93, doi:10.1016/j.drudis.2020.10.010.

20. Liebman, M. The Role of Artificial Intelligence in Drug Discovery and Development Chemistry International**,** 44 (2022**)** 16-19, doi:doi:10.1515/ci-2022-0105.
21. Alam, K.; Mazumder, A.; Sikdar, S.; Zhao, Y.M.; Hao, J.; Song, C.; Wang, Y.; Sarkar, R.; Islam, S.; Zhang, Y.; et al. Streptomyces: The biofactory of secondary metabolites Frontiers in Microbiology**,** 13 (2022**)**, doi:10.3389/fmicb.2022.968053.
22. Aulner, N.; Danckaert, A.; Ihm, J.; Shum, D.; Shorte, S.L. Next-Generation Phenotypic Screening in Early Drug Discovery for Infectious Diseases Trends in Parasitology**,** 35 (2019**)** 559-570, doi:10.1016/j.pt.2019.05.004.
23. Barbolla, I.; Hernández-Suárez, L.; Quevedo-Tumailli, V.; Nocedo-Mena, D.; Arrasate, S.; Dea-Ayuela, M.A.; González-Díaz, H.; Sotomayor, N.; Lete, E. Palladium-mediated synthesis and biological evaluation of C-10b substituted Dihydropyrrolo[1,2-b]isoquinolines as antileishmanial agents European Journal of Medicinal Chemistry**,** 220 (2021**)**, doi:10.1016/j.ejmech.2021.113458.
24. Herrera-Acevedo, C.; Perdomo-Madrigal, C.; Muratov, E.N.; Scotti, L.; Scotti, M.T. Discovery of Alternative Chemotherapy Options for Leishmaniasis through Computational Studies of Asteraceae ChemMedChem**,** 16 (2021**)** 1234-1245, doi:10.1002/cmdc.202000862.
25. Howl, J.; Jones, S. A new biology of cell penetrating peptides Peptide Science**,** 113 (2021**)**, doi:10.1002/pep2.24154.
26. Hu, R.S.; Hesham, A.E.L.; Zou, Q. Machine Learning and Its Applications for Protozoal Pathogens and Protozoal Infectious Diseases Frontiers in Cellular and Infection Microbiology**,** 12 (2022**)**, doi:10.3389/fcimb.2022.882995.
27. Jamal, S.; Scaria, V. Cheminformatic models based on machine learning for pyruvate kinase inhibitors of Leishmania mexicana BMC Bioinformatics**,** 14 (2013**)**, doi:10.1186/1471-2105-14-329.
28. Kwofie, S.K.; Broni, E.; Dankwa, B.; Enninful, K.S.; Kwarko, G.B.; Darko, L.; Durvasula, R.; Kempaiah, P.; Rathi, B.; Miller, W.A., III; et al. Outwitting an old neglected nemesis: A review on leveraging integrated data-driven approaches to aid in unraveling of leishmanicides of therapeutic potential Current Topics in Medicinal Chemistry**,** 20 (2020**)** 349-366, doi:10.2174/1568026620666200128160454.
29. Monteagudo, M.C.; González-Díaz, H. New experimental and computational tools for drug discovery: Medicinal chemistry, molecular docking, and machine learning - Part-VI Current Topics in Medicinal Chemistry**,** 18 (2018**)** 2325-2326, doi:10.2174/1568026619666181130122945.
30. Sánchez-Suárez, J.; Bernal, F.A.; Coy-Barrera, E. Colombian contributions fighting leishmaniasis: A systematic review on antileishmanials combined with chemoinformatics analysis Molecules**,** 25 (2020**)**, doi:10.3390/molecules25235704.
31. Soni, M.; Pratap, J.V. Development of Novel Anti-Leishmanials: The Case for Structure-Based Approaches Pathogens**,** 11 (2022**)**, doi:10.3390/pathogens11080950.
32. Teixeira, M.A.C.; Belloze, K.T.; Cavalcanti, M.C.; Silva-Junior, F.P. Data mart construction based on semantic annotation of scientific articles: A case study for the prioritization of drug targets Computer Methods and Programs in Biomedicine**,** 157 (2018**)** 225-235, doi:10.1016/j.cmpb.2018.01.010.
33. Castillo-Garit, J.; Flores-Balmaseda, N.; Alvarez, O.; Hai, P.-T.; Perez-Donate, V.; Torrens, F.; Perez-Gimenez, F. Computational Identification of Chemical Compounds with Potential Activity against Leishmania amazonensis using Nonlinear Machine Learning Techniques Current Topics in Medicinal Chemistry**,** 18 (2018**)** 2347-2354, doi:10.2174/1568026619666181130121558.
34. Clemente, C.M.; Robledo, S.M.; Ravetti, S. Menthol carbonates as potent antiparasitic agents: synthesis and in vitro studies along with computer-aided approaches Bmc Complementary Medicine and Therapies**,** 22 (2022**)**, doi:10.1186/s12906-022-03636-8.
35. Guan, L.; Yang, H.; Cai, Y.; Sun, L.; Di, P.; Li, W.; Liu, G.; Tang, Y. ADMET-score - a comprehensive scoring function for evaluation of chemical drug-likeness Medchemcomm**,** 10 (2019**)** 148-157, doi:10.1039/c8md00472b.

36. Halder, A.K.; Natalia Dias Soeiro Cordeiro, M. Advanced in Silico Methods for the Development of Anti-Leishmaniasis and Anti-Trypanosomiasis Agents Current medicinal chemistry**,** 27 (2020**)** 697-718, doi:10.2174/0929867325666181031093702.
37. Scotti, L.; Ishiki, H.; Mendonca Junior, F.J.B.; da Silva, M.S.; Scotti, M.T. Artificial Neural Network Methods Applied to Drug Discovery for Neglected Diseases Combinatorial Chemistry & High Throughput Screening**,** 18 (2015**)** 819-829, doi:10.2174/1386207318666150803141219.
38. Singh, N.; Shah, P.; Dwivedi, H.; Mishra, S.; Tripathi, R.; Sahasrabuddhe, A.A.; Siddiqi, M.I. Integrated machine learning, molecular docking and 3D-QSAR based approach for identification of potential inhibitors of trypanosomal N-myristoyltransferase Molecular Biosystems**,** 12 (2016**)** 3711-3723, doi:10.1039/c6mb00574h.
39. Singh, P.; Kumar, A. Deciphering the function of unknown Leishmania donovani cytosolic proteins using hyperparameter-tuned random forest Network Modeling and Analysis in Health Informatics and Bioinformatics**,** 9 (2019**)**, doi:10.1007/s13721-019-0208-2.