Revolutionizing Drug Discovery And Preclinical Research Via Artificial Intelligence: A Targeted Literature Review

Abdelghani I[1], Jdidi H[1], Boukhris Y[1], Louhichi KE[1], Roch B[2], François C[3], Toumi M[3], Bakhutashvili A[4]
[1]Creativ-Ceutical, Tunis, Tunisia, [2]Creativ-Ceutical, Paris, France, [3]Aix-Marseille University, Marseille, France, [4]MarcoPolo, Luxembourg

ISPOR US, Washington, DC
USA Area and Virtual
May 15-18, 2022

MSR8

Background and objective

- Artificial intelligence (AI) is defined as “machines that mimic cognitive functions that humans associate with the human mind, such as learning and problem solving”.
- Many notable signs of progresses have occurred lately in this field. From drug discovery to clinical trials, AI has rapidly taken place, investing the different sectors distinctly.
- AI can be applied to improve efficiency, reduce time, avoid animal ethical issues, and defects in the pre-clinical phase which would otherwise lead to some level of failure at the time of clinical development or to a delay in a detection of effective compounds and future drugs.
- This study aims to provide an overview of the use of computed technologies in drug discovery and pre-clinical development.
- The article “Artificial Intelligence in drug discovery and development” by Paul et al.[1] was used as the primary reference.

Methodology

- A bibliographic search was performed in OVID (MEDLINE + EMBASE) based on search strategy showed in Table 1.
- This search was restricted to studies published between January 01, 2020 and December 01, 2021, to better reflect the latest evolution of the domain and due to the high number of references retrieved.
- For these last two years, 810 titles were retrieved from which 197 were out of scope.
- The criteria for articles selection were publications that were available in full-text open-access, and in the AMU library. Selected articles had to include detailed description of AI methods, clear identification of the name of AI tools used with or without links to AI tools. In total, 40 articles were selected.
- Information on the description of the tasks performed using AI, the tools, the assessment of the performance and limitations of existing tools, the explored database, and the expected future developments were extracted.

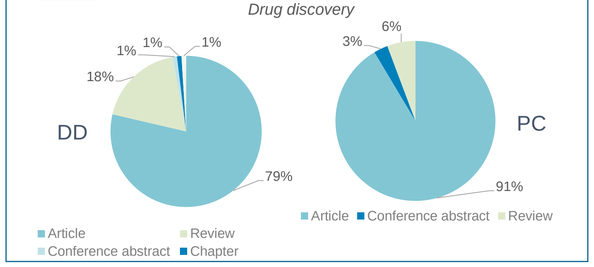
Table 1: Search strategy keywords

Drug discovery and preclinical search terms	AI studies search terms
QSAR	Artificial intelligence
Docking experiments, molecular docking programs	Machine learning, deep learning, reinforcement learning
Holistic simultaneous in silico models	Learning algorithm$
Virtual screening / virtual combinatorial chemistry space	Neural networks
In-silico biomimetic model	Natural language processing
In-silico drug absorption	Expert system$, intelligent agent$
Organ-on-a-chip technology	Supervised learning, unsupervised learning
ANTARES	
ReThink3R	

Results

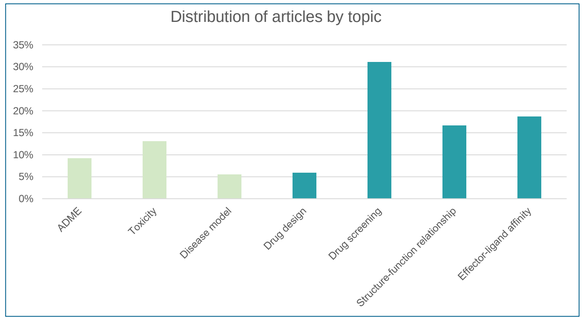
- In total, the 613 publications abstracts that were analyzed, were mostly from the US and China.
- 75% of them were predominantly related to drug discovery and classified as drug design, drug screening, structure-function relationship, and effector-ligand affinity.
- On the other hand, the number of publications in preclinical research was roughly ⅓ of those in drug discovery. They were classified under ADME (Administration /Distribution /Metabolism/ Excretion) prediction, toxicity or disease models. Publications on disease modelling were sparse and not as informative as those in drug discovery or in ADME (Figure 2).
- In general, AI as a proxy to in vivo studies appeared under-represented. This may be due to either the difficulty to model biological systems or because of a bibliographic search bias related to the selected OVID databases or key words choice («3D», which refers to in-silico studies, was not included, for instance).

Figure 1: *Distribution of selected by type of publication; PC: preclinical; DD: Drug discovery*



- Articles on the use of AI in COVID-19 studies were numerous, reflecting the priority to adapt AI and its importance during the pandemic. However, to avoid a representation bias, we selected only a few full-text articles on COVID-19 for the analysis (the most illustrative and using different approaches).
- The AI tools that were most often used are Autodock (and its different subtools such as Autodock 4 and Autodock Vina), R package, DeepCoy, Discovery Studio, RDKit, and the Python library.

Figure 2: *Distribution of selected articles by category: green: preclinical topics; blue: drug discovery topics*



- Examples of databases consulted are ChEMBL, BindingDB, and ZINC database.
- The BIOiSIM Platform was an example of an attempt at an in-silico organ (skin).[2] Another platform is freely available for the in-silico screening and design of new anticancer peptide chemotherapeutics at https://research.timmons.eu/ennaact.[3]
- An example of a preclinical AI application was shown in an article about simulating a rat liver microsome which provided the link to the available prediction models: https ://opendata.ncats .nih.gov/adme.[4] Pharmaceutical companies also develop their own platforms such as the Bayer’s ADMET platform, including a chemical library and AI tools: CypScore and MetScore.
- The most frequent sponsors of AI in the pharmaceutical industry were Bayer, Sanofi, AstraZeneca, and Japanese companies.

CONCLUSIONS

- Using mathematical modelling and feature extraction, machine learning, and neural networks have optimized, accelerated, and significantly reconfigured non-clinical research procedures.
- AI enables researchers to explore novel drug lead compounds what would have been impossible without this progress. Therefore, AI techniques yield candidates with optimal efficacy and safety through predicting molecular attributes in-silico from their directed structure and examining disease models.
- The other side of the coin is that the computed tools still face a few challenges:
  - Their performance is affected by the scarcity and high cost of reliable high-quality databases. This may impact on the achievement of good quality descriptors required for reaching a performant and reproducible digital molecular triage.
  - Even though algorithms may have improved the minimization of errors and system instability, the accuracy of the data gathered may be less than expected due to non-reproducibility of results and the risk of over/underfitting. These issues can be addressed by expanding the sample size and using cross-validation.
  - Even though algorithms existed for a long time, their biological findings are still novel. So, documented follow-up and returns of experience are missing.

REFERENCES
1. Paul D et al. Drug Discov Today. 2021; 26:80-93
2. Maharao N et al. Drug Des Devel Ther. 2020;14:2307-2317
3. Timmons PB, Hewage CM. Biomed Pharmacother. 2021;133:111051.
4. Siramshetty V et al. SLAS Discov. 2021; 26:1326-1336