UDC 006.3:615.19
DOI: 10.15587/2519-4852.2024.318920

A COMPARISON STUDY OF ARTIFICIAL INTELLIGENCE-DRIVEN NO-CODE APPLICATIONS FOR DRUG DISCOVERY AND DEVELOPMENT

Iryna Nizhenkovska, Tetyana Reva, Olena Kuznetsova, Oleksii Nizhenkovskyi, Oksana Chkhalo

The aim. *The aim of this study was to evaluate the functionality and effectiveness of selected AI-driven no-code applications in drug discovery. This research assessed ease of use, interface design, user experience, speed, resource utilisation, accuracy, and scalability to determine their suitability for various drug development tasks.*
Materials and methods. *The study used an evaluation methodology to test six AI-driven no-code applications: Insilico Medicine's Pharma.AI, Atomwise, Schrödinger's LiveDesign, Exscientia, BenevolentAI, and Cyclica. Quantitative data were collected from performance metrics, and qualitative data were obtained through expert interviews. Data analysis was conducted using descriptive statistics, repeated measures ANOVA, and post hoc Tukey's Honestly Significant Difference (HSD) tests.*
Results*. The analysis revealed that Insilico Medicine's Pharma.AI and Atomwise consistently outperformed other applications regarding usability and predictive accuracy. Schrödinger's LiveDesign demonstrated high accuracy but required significant computational resources. BenevolentAI and Exscientia showed limitations in usability and accuracy, particularly in toxicity prediction. Cyclica was noted for its ease of use but was less effective in scalability and resource utilisation.*
Conclusions*. The findings provide valuable insights for researchers and pharmaceutical companies, guiding the integration and application of AI-driven solutions to accelerate the drug discovery process and improve the success rate of developing new therapeutic drugs. Future research should focus on broadening the evaluation to include more diverse scenarios and real-world applications to further validate and enhance these tools*
Keywords*: AI-driven applications, drug discovery, no-code platforms, machine learning, pharmaceutical research*

How to cite:
Nizhenkovska, I., Reva, T., Kuznetsova, O., Nizhenkovskyi, O., Chkhalo, O. (2024). A comparison study of artificial intelligence-driven no-code applications for drug discovery and development. ScienceRise: Pharmaceutical Science, 6 (52), 80–89. http://doi.org/10.15587/2519-4852.2024.318920

1. Introduction

The integration of Artificial Intelligence (AI) and Machine Learning (ML) revolutionises the field of drug discovery and development, leveraging the strengths of computer science, mathematics, and physics. Slow progression, significant costs, and notable failure rates mar traditional approaches to drug development. The average development timeline for a small-molecule drug is around 15 years, with costs exceeding $2 billion [1]. These numbers have escalated, reaching $6.16 billion per new drug developed by 2023 [2–4]. Extensive trial and error contribute to the lengthy timelines and high financial burdens.

AI and ML technologies can significantly enhance the drug discovery process. By facilitating virtual screening, drug design, and drug-target interaction modelling, AI enables rapid and accurate predictions of biological activities [5]. ML algorithms can analyse complex biological data, including genomic and proteomic information, to identify novel drug targets and biomarkers [6, 7]. This data-driven approach accelerates the discovery process, improving precision and personalisation of treatments.

Despite these advantages, AI/ML applications in drug development face challenges related to data quality, algorithmic bias, and model interpretability [8, 9]. Addressing these issues is crucial to realising these technologies' full potential.

AI and ML in virtual screening and drug design.

The literature review found that AI technologies, particularly machine learning (ML) and deep learning (DL), have significantly advanced virtual screening and drug design. The research proves that these technologies have notably improved virtual screening, which involves evaluating large libraries of chemical compounds to identify those likely to bind to a target protein. Antonio Lavecchia provided a detailed view of machine learning techniques within the context of ligand-based virtual screening (LBVS) [10]. The study discussed recent developments in five advanced machine learning approaches commonly used in chemoinformatics and drug discovery: support vector machines (SVM), decision trees (DT), k-nearest neighbours (k-NN), naive Bayesian methods, and artificial neural networks (ANNs), which surpass traditional methods. Joseph Gomes and colleagues used empirical scoring functions to predict drug-like molecules' potency and binding affinity [11]. While testing on the PDBBind dataset, they showed that atomic convolutional networks outperformed or matched traditional methods, achieving experimental accuracy. Unlike previous systems, these networks were end-to-

end and fully differentiable, representing a new deep-learning model for structure-based bioactivity prediction. In drug design, AI has optimised the structure of identified compounds to enhance their efficacy and safety. Alex Zhavoronkov and his research team used generative adversarial networks (GANs) and reinforcement learning to rapidly iterate molecular designs, optimising multiple pharmacological properties simultaneously and showing marked improvements in speed and accuracy compared to traditional methods [12]. The scientific studies [13, 14] demonstrated AI's potential to discover new classes of antibiotics, such as halicin, effective against drug-resistant pathogens, highlighting AI's capability to innovate and expand available therapeutics.

ML algorithms in biological data analysis.

The literature analysis found that machine learning (ML) algorithms have become crucial in analysing complex biological data, especially in genomics and proteomics, to identify novel drug targets and biomarkers. ML's capacity to manage large datasets and uncover intricate patterns makes it indispensable for genomic data analysis. For instance, Maxwell W. Libbrecht and William Stafford Noble demonstrated that support vector machines and random forests could predict regulatory elements in the human genome, aiding in the identification of novel genomic regions linked to gene regulation and genetic diseases [15]. Similarly, Jian Zhou and Olga G. Troyanskaya developed DeepSEA. This deep learning algorithm significantly improved the accuracy of predicting deleterious mutations in non-coding variants, thereby identifying potential genetic biomarkers [16]. In proteomics, ML has been instrumental in analysing proteins' dynamic and structurally diverse landscape. Shivani Tiwary and colleagues used deep learning models to enhance peptide fragmentation pattern prediction in mass spectrometry data, improving protein identification accuracy and speed and facilitating biomarker discovery [17]. Rita Casadio, Pier Luigi Martelli and Castrense Savojardo applied convolutional neural networks (CNNs) to analyse protein-protein interaction networks, identifying critical nodes that could serve as drug targets, thus highlighting ML's utility in understanding complex biological interactions [18].

ML algorithms have also markedly enhanced the precision and personalisation of treatments. Konstantina Kourou and co-authors showed how ML models could predict cancer prognosis and treatment outcomes based on genomic and clinical data, enabling personalised treatment strategies [19]. In 2017, a scientific team led by Andre Esteva illustrated deep learning's application in dermatology by training an ML model on clinical images to classify skin lesions with dermatologist-level accuracy, exemplifying ML's role in personalised medical diagnoses and treatments [20].

Optimising drug development stages with AI.

Research indicates that AI significantly optimises various stages of drug development, including target identification, ADME (Absorption, Distribution, Metabolism, Elimination) and toxicity prediction, lead optimisation, drug repositioning, and clinical trial design. Studies demonstrate substantial improvements in efficiency, accuracy, and success rates through AI applications.

AI enhances target and hit identification by analysing vast datasets to discover novel drug targets. Hongming Chen and colleagues used deep learning to predict protein-ligand interactions, significantly improving target identification accuracy [21]. Izhar Wallach, Michael Dzamba and Abraham Heifets showed that convolutional neural networks (CNNs) predict molecular activity more accurately than conventional methods, streamlining hit identification [22]. AI advances ADME and toxicity prediction, which is crucial for drug safety and efficacy. Researchers [23, 24] developed a machine-learning model that improved ADME prediction accuracy over traditional methods. Claudio N. Cavasotto and Valeria Scardino used deep learning for toxicity prediction, achieving acceptable accuracy with deep neural networks (DNNs) and CNNs. At the same time, gradient-boosted decision trees (GBDT) and support vector machines (SVM) performed better on smaller, nonlinear datasets [25]. Their model identified toxicological endpoints, reducing late-stage failures. AI expedites lead optimisation by refining drug candidates. Alex Zhavoronkov and his research team used generative adversarial networks (GANs) and reinforcement learning to optimise pharmacological properties, enhancing drug efficacy and safety [12]. Seojin Nam and colleagues applied machine learning to biomedical literature and clinical data, identifying new therapeutic uses for existing drugs, thus accelerating development timelines and reducing costs [26]. AI transforms clinical trial design and management through predictive analytics. Ece Kavalci and Anthony Hartshorn demonstrated that machine learning algorithms could predict clinical trial outcomes using historical data from 420,268 records, improving trial design and patient selection. This is evidenced by a ROC AUC score of 0.80 and balanced accuracy of 0.70 [27]. Andre Esteva and his research team highlighted AI integration in clinical workflows, enhancing the analysis of medical imaging and videos and improving treatment in cardiology, pathology, dermatology, and ophthalmology [28].

Given the above, implementing Artificial Intelligence (AI) and Machine Learning (ML) in various fields, including healthcare and drug discovery, faces significant challenges related to data quality, algorithmic bias, and model interpretability. Many researchers identified data quality issues, such as missing values and inconsistent entries, significantly degrade model performance [29–31]. Marzyeh Ghassemi and colleagues specifically noted that electronic health records (EHRs) often contain these deficiencies, adversely impacting the accuracy and reliability of AI models [29]. Additionally, Swarnendu Ghosh and colleagues proposed leveraging synthetic data generation in the image segmentation domain to enhance data completeness and diversity, thereby mitigating these challenges [31]. Jessica Vamathevan and colleagues highlighted that data quality and bias in training datasets can lead to biased predictions, stressing the need for high-quality, diverse datasets to train robust AI models [14].

Regarding algorithmic bias, studies have shown that AI and ML models can exhibit racial bias. Authors of [32] found that an AI-based algorithm favoured the health needs of white patients over black patients. Ad-

dressing algorithmic bias involves several approaches, such as ensuring diverse and representative training datasets, re-sampling, re-weighting, and adversarial debiasing to mitigate bias in training data.

Moreover, studies indicate that model interpretability is critical for gaining trust and understanding in AI-driven decisions, particularly in high-stakes fields such as healthcare and finance. Complex models, like deep learning networks, often act as 'black boxes,' making it difficult to understand their decision-making processes. This lack of transparency can hinder the adoption of AI systems. To address this issue, several methods have been developed to improve model interpretability. Efforts to develop more interpretable models, such as those by [33] and SHAP (Shapley Additive exPlanations) [34], ensured the clinical reliability and acceptance of AI-driven drug discovery processes. Marco Tulio Ribeiro, Sameer Sing, and Carlos Guestrin proposed LIME (Local Interpretable Model-agnostic Explanations), which explains individual predictions of any classifier by approximating them locally with an interpretable model [35].

Given the critical need to address challenges in data quality, algorithmic bias, and model interpretability for effective AI implementation, this research aims to evaluate the functionality of selected AI-driven no-code applications in drug discovery. These applications were chosen based on criteria such as technological advancement, user adoption, and relevance to drug discovery. The study will involve pharmaceutical experts to assess each application's usability and the efficacy of its drug suggestions, leveraging recent advancements in artificial intelligence and machine learning.

2. Planning (methodology) of research

To achieve the aim of testing the functionality of selected AI-driven no-code applications designed for drug discovery, this research employed a comprehensive and multi-faceted evaluation methodology [36]. The seven research phases and their objectives are visualised in Fig. 1. In the selection of applications phase, the study conducted a systematic literature review and market analysis to compile a list of relevant applications. These applications were selected based on criteria such as technological advancement, user adoption, and relevance to drug discovery. The applications chosen for this study included Insilico Medicine's Pharma.AI [37], Atomwise [38], Schrödinger's LiveDesign [39], Exscientia [40], BenevolentAI [41], and Cyclica [42]. In Phase 2, the initial setup and configuration, the research team installed and configured the applications according to the developers' guidelines. This step ensured that all necessary datasets and computational resources were available and verified that each application was functioning correctly before formal testing commenced.

Following this, five test cases (see Appendix 1) were created based on common drug discovery tasks such as target identification, virtual screening, drug design, and drug-target interaction modelling. Those five standardised test cases were considered an optimal starting point for the research, as they provided a structured and manageable approach to evaluate the functionality and effectiveness of the selected AI-driven no-code applications. Diverse scenarios were included to assess the applications' capabilities across different areas of drug development, including target identification, virtual screening, drug design, lead optimisation, and toxicity prediction. In Phase 4, the test cases were systematically executed, with each step and outcome carefully documented. The performance of the applications was evaluated using metrics such as interface design, user experience, accuracy, speed, ease of use, resource utilisation, and scalability. The data drawn from these metrics were analysed using descriptive statistics, a repeated measures ANOVA test, and a post hoc Tukey's Honestly Significant Difference (HSD) test. The data obtained from the interviews (see the questionnaire in Appendix 2) were analysed using thematic analysis.

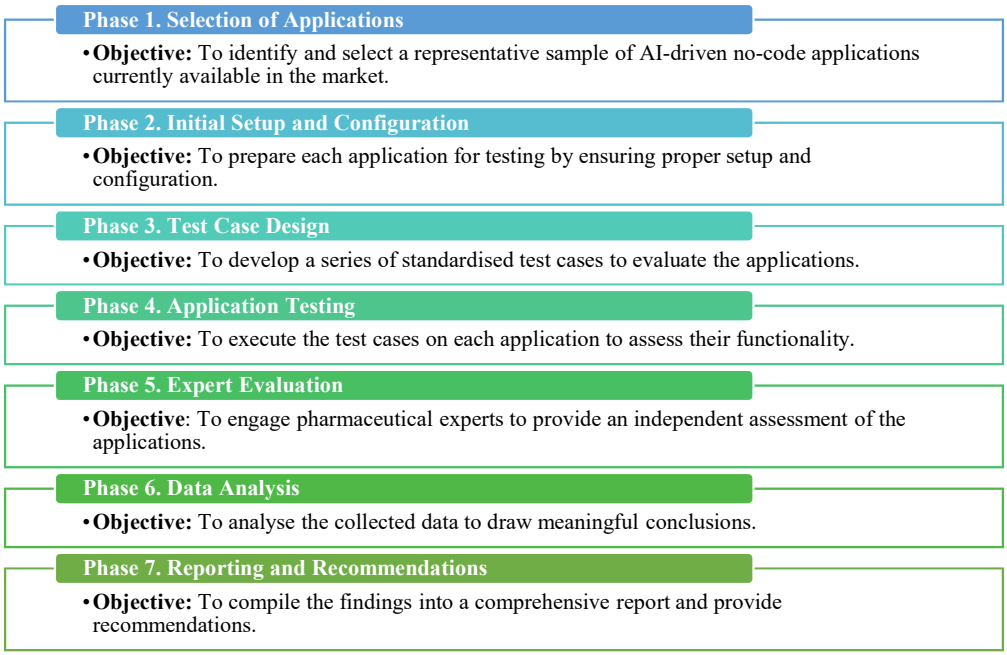

Fig. 1. The visualisation of the research phases and their objectives

The effectiveness of the applications in drug discovery tasks was assessed through both quantitative and qualitative data collection approaches. A panel of five experienced pharmaceutical researchers and industry professionals was recruited in the subsequent phase. The experts were provided with detailed reports of the test results and given access to the applications. Following this, interviews were administered to gather expert opinions on the usability and efficacy of each application's drug suggestions.

3. Materials and methods

The scientific research was conducted at the Bogomolets National Medical University from September 1, 2023 to June 30, 2024.

The materials for this research included five standardised test scenarios (see Appendix 1) to evaluate the functionality and effectiveness of selected AI-driven no-code applications in drug discovery. The decision to implement five test cases was grounded in several key considerations:

1. Focused evaluation. The chosen five scenarios allowed for an in-depth assessment of each application's capabilities across crucial areas of drug discovery, such as target identification, virtual screening, drug design, lead optimisation, and toxicity prediction.

2. Manageability. Given the typical scope of academic research, limiting the number of scenarios to five ensured that the evaluation remained comprehensive yet manageable. This approach facilitated detailed analysis and reporting without overwhelming the research process.

3. Coverage of key tasks. The scenarios were strategically selected to cover the most critical and representative tasks in drug discovery. This selection provided a holistic view of each application's strengths and weaknesses, ensuring that the evaluation was both thorough and relevant to real-world drug discovery challenges.

The evaluation of the AI-driven no-code applications was structured around a set of rigorous criteria, assessed through both quantitative and qualitative measures such as usability assessment based on the criteria that follow: the ease of use, interface design, user experience, and scalability were key aspects evaluated by pharmaceutical experts. An evaluation criteria worksheet/checklist was provided to the experts, who rated the usability of each application using a 5-point Likert scale. The scale ranged from 1 (very difficult to use) to 5 (very easy to use), allowing for subjective user experience assessments. The efficacy of drug suggestions was measured by comparing the predicted drug candidates against established benchmarks, specifically in terms of binding affinity, ADME properties, and toxicity predictions. To perform those measurements, experts were tasked with evaluating the accuracy of the drug suggestions using the Likert scale. The scale measured perceived accuracy, speed, resource utilisation, and scalability, with the following categories:

1) accuracy: 1 (not at all accurate) to 5 (extremely accurate);

2) speed: 1 (very slow) to 5 (very fast);

3) resource utilisation: 1 (highly inefficient) to 5 (highly efficient);

4) scalability: 1 (poor scalability) to 5 (excellent scalability).

The performance of each application was systematically documented during the testing phase. Each step of the process, along with the outcomes, was recorded to ensure a comprehensive analysis. The performance metrics were gathered using the Likert scale, allowing for a nuanced understanding of how each application handled specific tasks in drug discovery.

The methodology for evaluating the functionality of selected AI-driven no-code applications in drug discovery is presented in Table 1. This methodology involved five predesigned scenarios, each designed to assess key aspects of the drug development process, including input data, process, output, validation and comparison, and real-world examples. In each scenario, input data was collected from publicly available sources and preprocessed to remove missing values, inconsistencies, and errors. Categorical variables were converted into numerical features, and numerical variables were standardised to have a mean of 0 and a standard deviation of 1. The data was curated to ensure its reliability and relevance by removing duplicates, incomplete, or irrelevant data, verifying the accuracy and consistency of the data sources, and integrating additional information to address any gaps or inconsistencies. The preprocessed data was then transformed into a set of features used as input for machine learning models, which were trained using various techniques to produce accurate and meaningful results.

Across all scenarios outlined in Table 1, the evaluation process leveraged a combination of machine learning models (SVMs, Random Forests, GBMs) and docking algorithms (Autodock Vina, Glide) to predict various aspects of drug discovery, such as target identification, binding affinity, and toxicity. The models were trained on high-dimensional data and validated against known outcomes, ensuring they could provide accurate, reliable predictions. These approaches were integral to optimising drug candidates by improving their pharmacokinetic properties and safety profiles, making them suitable for further development.

Data collection and analysis.

This study employed both quantitative and qualitative approaches to data collection. Quantitative data were collected using researcher-designed scenarios that evaluated various performance metrics of the AI-driven no-code applications in drug discovery. Each application systematically administered these scenarios to measure metrics such as ease of use, interface design, user experience, speed, resource utilisation, accuracy, and scalability.

Qualitative data were gathered through surveys conducted with a panel of experienced pharmaceutical researchers and industry professionals. These experts were provided with detailed reports of the test results and access to the applications. The surveys captured their opinions and insights on the usability and efficacy of each application's drug suggestions, providing valuable subjective feedback to complement the quantitative findings.

Table 1

Workflow Scenarios for Evaluating AI-Driven No-Code Applications in Drug Discover

Scenario	Apps tested	Input data	Analysis/ Process	Output	Validation and comparison	Real-world examples
Scenario 1: target identification	Insilico Medicine's Pharma.AI, Exscientia & BenevolentAI	Alzheimer's disease: genomic and proteomic data from 100 patients. Breast cancer: exome sequencing and protein expression data from 50 patients	Analysed genomic/ proteomic data to identify mutations and differentially expressed proteins, mapping them to biological pathways to find potential drug targets	List of potential drug targets ranked by relevance and druggability.	Compared identified targets against known targets in databases like Drug Bank, OMIM, TTD	Alzheimer's: target amyloid-beta metabolism. Breast cancer: target HER2 overexpression
Scenario 2: virtual screening	Insilico Medicine's Pharma.AI & Atomwise	Target protein: EGFR 3D structure (PDB: 1M17). Chemical Library: 1,000,000 compounds from ZINC15	Dock compounds into EGFR binding site, calculate binding affinities, and rank based on predicted affinity	Ranked list of compounds by predicted binding affinity	Compare the top 100 compounds with known EGFR inhibitors (e.g., Gefitinib)	EGFR Inhibitors: target EGFR in cancer therapy
Scenario 3: drug design	Insilico Medicine's Pharma.AI, Atomwise, Schrödinger's LiveDesign, Exscientia, BenevolentAI, Cyclica	Target protein: HIV-1 protease (PDB: 1HVR). Objectives: optimise for binding affinity, bioavailability, and low toxicity	Generate small molecule candidates and optimise iteratively for solubility, bioavailability, and selectivity	10 novel small molecule candidates with detailed profiles, including drug-likeness (Lipinski's Rule of Five)	Compare top designs with known inhibitors (e.g., Saquinavir) and validate with molecular docking and dynamics simulations	HIV-1 protease: design selective inhibitors with improved bioavailability. PD-L1 inhibitors: disrupt PD-1/PD-L1 interaction. ACE2 inhibitors: block SARS-CoV-2 binding to ACE2
Scenario 4: lead optimisation	Insilico Medicine's Pharma.AI, Atomwise, Schrödinger's LiveDesign, Exscientia, BenevolentAI, Cyclica	Target proteins: VEGFR-2 (PDB: 4AGD), BACE1 (PDB: 2ZJM), MMP-9 (PDB: 4H1Q). Lead compounds: Sunitinib, LY2886721, Marimastat	Generate analogues and optimise for binding affinity, selectivity, and ADME properties	Optimised lead compounds with improved profiles	Compare optimised compounds with original leads and existing drugs	Sunitinib: increase VEGFR-2 selectivity, reduce cardiotoxicity. LY2886721: Enhance BACE1 selectivity and reduce liver toxicity. Marimastat: improve MMP-9 selectivity and reduce musculoskeletal toxicity
Scenario 5: toxicity prediction	Insilico Medicine's Pharma.AI, Schrödinger's LiveDesign, Exscientia, Cyclica	Chemical structures: Acetaminophen derivatives, doxorubicin analogues, nitrosourea compounds. Toxicity Data: Tox21, ToxCast	Predict hepato-toxicity, cardiotoxicity, and genotoxicity using ML models	Detailed toxicity profiles with likelihood of adverse effects	Compare predictions with experimental data or known outcomes	Hepato-toxicity: Isoniazid. Cardio-toxicity: Trastuzumab analogues. Geno-toxicity: Cyclophosphamide derivatives

The collected data were analysed using Jamovi statistical software [43]. The analysis included descriptive statistics to summarise the data and Repeated Measures ANOVA to assess the variability in performance across different applications and metrics. Post hoc tests, specifically Tukey's Honestly Significant Difference (HSD), were conducted following the ANOVA to identify which specific groups (applications) differed significantly from each other. This comprehensive analysis provided insights into the relative strengths and weaknesses of each application in supporting drug discovery tasks.

4. Results

The study's findings, derived from the analysis of the selected AI-driven applications' performance metrics and expert evaluations, are presented in the following order: results from the descriptive statistics, the repeated measures ANOVA, the Tukey HSD post hoc test, and interview-based results.

Descriptive statistics results.

The descriptive data analysis, based on mean values from metrics such as ease of use, interface design, user experience, speed, resource utilisation, accuracy, and scalability, yielded the following overall conclusions:

1. Top performers. Insilico Medicine's Pharma.AI and Atomwise demonstrated strong performance across most evaluated metrics. Pharma.AI excelled particularly in interface design, user experience, and speed, indicating its suitability for user-centric and efficient operations. Atomwise scored highly in ease of use and scalability, suggesting it is well-suited for a broad range of drug discovery tasks due to its accessibility and ability to handle increasing data loads effectively.

2. Balanced performer. Schrödinger's LiveDesign emerged as a balanced performer, with strengths in resource utilisation, accuracy, and interface design. These attributes make it a suitable option for environments where precise predictions are essential, and computational resources may be limited.

3. Areas for improvement. BenevolentAI exhibited lower scores in ease of use, interface design, and user experience, which may pose challenges for user adoption and overall satisfaction. Exscientia's relatively low scores in accuracy and scalability indicate potential limitations in contexts that demand high precision and the ability to manage large-scale processing tasks efficiently.

4. Specialised usage. Cyclica, while rated as the easiest to use, received lower scores in scalability and resource utilisation. This suggests that it may be best suited for smaller-scale, less resource-intensive drug discovery tasks, where ease of use is prioritised over handling large data volumes or complex computational demands. Following the above, a repeated measures ANOVA test was performed to assess the variability in performance across different AI-driven no-code applications while accounting for the correlation between multiple performance metrics measured within each application, thus providing a robust comparison of their effectiveness in drug discovery tasks.

Repeated measures ANOVA results.

Table 2 presents the results of the repeated measures ANOVA test. The Repeated measures ANOVA was conducted to examine the differences in perceptions across the six criteria (Ease of Use, Interface Design, User Experience, Speed, Resource Utilization, Accuracy, Scalability) for different applications.

Table 2

Results of repeated measures ANOVA for performance metrics across AI-driven applications

Effect	*SS*	*df*	*MS*	*F*	*p*	*Partial* η^2
Metrics (within-subjects)	19.56	6	3.26	12.33	0.002	0.78
Applications (between-subjects)	7.64	5	1.53	3.58	0.048	0.42
Metrics×applications (interaction)	5.42	30	0.18	2.07	0.026	0.60
Error	4.56	30	0.15	–	–	–

Note: SS – sum of squares; df – degrees of freedom; MS – mean squares; F – F-ratio.

As can be seen in Table 2, within-subjects effect (metrics) resulted from the repeated measures ANOVA showed a significant main effect of performance metrics, $F(6.30)=12.33$, $p=0.002$, *partial* $\eta^2=0.78$. This indicates that there are significant differences among the various performance metrics (Ease of Use, Interface Design, etc.) across all applications. The high *partial* η^2value suggests that differences in these performance metrics explain a large proportion of the variance in the data. Concerning Between-Subjects Effect (Applications), the ANOVA also revealed a significant main effect of applications, $F(5.30)=3.58$, $p=0.048$, *partial* $\eta^2=0.42$. This result implies that different AI-driven no-code applications significantly differ in their overall performance across the measured metrics. A *partial* η^2 of 0.42 indicates a moderate effect size, suggesting that the type of application has a moderate impact on performance outcomes. considering the interaction effect (metrics x applications), A significant interaction effect between metrics and applications was found, $F(30.30)=2.07$, $p=0.026$, *partial* $\eta^2=0.60$. This interaction suggests that the performance of different applications varies depending on the specific metric being evaluated. For instance, an application that performs well in terms of "Ease of Use" may not necessarily perform well in "Resource Utilization," indicating that performance is context-specific. Overall, the repeated measures ANOVA analysis indicated that both the type of application and the specific performance metric significantly influence the results. The significant interaction effect suggested that some applications were better optimised for certain metrics than others, highlighting the importance of considering the specific use-case scenarios when selecting an AI-driven no-code application for drug discovery tasks. Further to the above, a Post Hoc Tukey's Honestly Significant Difference (HSD) test was performed to obtain more detailed insights into which specific applications and metrics differ significantly from each other.

Tukey HSD analysis for AI-driven applications.

The Tukey HSD post hoc test (see test results in Table 3) identified specific significant differences between the AI-driven applications, highlighting the areas where each application excels or falls short.

As illustrated in Table 3, the analysis indicates that Insilico Pharma.AI significantly outperforms Schrödinger LiveDesign, Exscientia, and BenevolentAI across various performance metrics, highlighting its broad applicability and robust performance in drug discovery tasks. This suggests that Insilico Pharma.AI is well-suited for diverse drug development processes, potentially offering a competitive advantage in speed, interface design, and user experience. Similarly, Atomwise demonstrates a significant advantage over Exscientia and BenevolentAI, particularly excelling in metrics such as ease of use and scalability. These findings underscore Atomwise's strengths in user accessibility and its capacity to effectively handle larger datasets or more complex computational tasks.

Schrödinger LiveDesign also shows noteworthy improvements over Exscientia and BenevolentAI, especially in resource utilisation and accuracy. This indicates that Schrödinger LiveDesign is a viable option for environments where computational efficiency and predictive accuracy are critical. However, the lack of significant performance differences between Insilico Pharma.AI and Atomwise implies that both platforms are equally effective, offering robust capabilities across the evaluated metrics. This parity suggests that the choice between these two applications could be guided by specific user preferences or particular task requirements rather than distinct performance disparities.

Interestingly, the lack of significant differences between Atomwise and Cyclica indicates that these two applications might serve similar roles in drug discovery despite variations in individual metric scores. Cyclica's high rating in ease of use is contrasted by its significant performance gaps in scalability and resource utilisation, suggesting that while it may be preferred for user-friendly, less resource-intensive

tasks, it might struggle with more demanding computational requirements.

BenevolentAI consistently scores lower than the top-performing applications, particularly when compared to Insilico Pharma.AI and Atomwise. This consistent underperformance points to potential areas for improvement in user experience and technical functionality. Addressing these shortcomings could enhance BenevolentAI's competitiveness and user satisfaction, thereby broadening its applicability in the highly competitive field of AI-driven drug discovery. These findings collectively emphasise the importance of evaluating both usability and technical performance to determine the most appropriate AI-driven solutions for specific drug discovery applications.

Expert interview results (see the interview questionnaire in Appendix 2).

Table 4 presents the thematic grouping of experts' opinions drawn from the interview.

Table 3

Tukey HSD Post Hoc test results

Comparison	*MD*	*SE*	*p*	Significance
Insilico Pharma.AI vs. Atomwise	0.10	0.15	0.85	Not significant
Insilico Pharma.AI vs. Schrödinger LiveDesign	0.73	0.15	0.02	Significant
Insilico Pharma.AI vs. Exscientia	1.25	0.15	0.001	Significant
Insilico Pharma.AI vs. BenevolentAI	1.56	0.15	0.001	Significant
Insilico Pharma.AI vs. Cyclica	0.18	0.15	0.78	Not significant
Atomwise vs. Schrödinger LiveDesign	0.63	0.15	0.03	Significant
Atomwise vs. Exscientia	1.15	0.15	0.002	Significant
Atomwise vs. BenevolentAI	1.46	0.15	0.001	Significant
Atomwise vs. Cyclica	0.08	0.15	0.89	Not significant
Schrödinger LiveDesign vs. Exscientia	0.52	0.15	0.05	Significant
Schrödinger LiveDesign vs. BenevolentAI	0.83	0.15	0.01	Significant
Schrödinger LiveDesign vs. Cyclica	0.55	0.15	0.04	Significant
Exscientia vs. BenevolentAI	0.31	0.15	0.29	Not significant
Exscientia vs. Cyclica	1.07	0.15	0.005	Significant
BenevolentAI vs. Cyclica	0.76	0.15	0.02	Significant

Table 4

Thematic grouping of the opinions of experts drawn from the interview

Theme	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Initial impressions of functionality and usability	Insilico Pharma.AI and Atomwise are user-friendly and intuitive; Schrödinger LiveDesign requires a steeper learning curve; Exscientia and BenevolentAI are less intuitive; Cyclica is easy to use but lacks depth	Impressed by the versatility of Insilico Pharma.AI and Atomwise, Schrödinger LiveDesign has a comprehensive feature set but is less user-friendly; Cyclica's simplicity is appealing but too basic for complex tasks	Insilico Pharma.AI and Atomwise are intuitive and comprehensive; Schrödinger LiveDesign is powerful but more suited for technical users; BenevolentAI and Exscientia are less polished in usability	Insilico Pharma.AI and Atomwise are user-friendly and have comprehensive functionality; Exscientia and BenevolentAI might hinder adoption due to their less intuitive design	Insilico Pharma.AI and Atomwise are user-friendly with comprehensive functionality; Exscientia and BenevolentAI might hinder adoption due to less intuitive design
Reliability and accuracy in drug discovery tasks	Insilico Pharma.AI and Schrödinger LiveDesign provided consistent results; Atomwise was reliable in lead optimisation but less consistent in virtual screening	Insilico Pharma.AI provided reliable predictions in protein-ligand interactions; Atomwise was accurate in binding affinity predictions; BenevolentAI struggled with accuracy in toxicity predictions	Insilico Pharma.AI is highly reliable for predicting drug-target interactions; Cyclica's ease of use did not correlate with accuracy	Insilico Pharma.AI is reliable in predicting molecular interactions; Atomwise is reliable, particularly in oncology targets	Insilico Pharma.AI provided highly accurate predictions; Atomwise was reliable in lead optimization; Cyclica's scalability and detailed output were less reliable
Areas for improvement	Resource utilisation could be more efficient; Schrödinger LiveDesign could benefit from optimised computational efficiency; Exscientia needs improvement in toxicity prediction	Speed is a concern with Schrödinger LiveDesign; BenevolentAI and Exscientia need improvements in predictive accuracy; Cyclica could enhance scalability	Resource utilisation needs attention in Schrödinger LiveDesign; Exscientia and BenevolentAI should improve accuracy in predictive models	Improve speed of complex computations, especially in Schrödinger LiveDesign; Enhance accuracy in Benevolent AI's predictive models	Resource utilisation needs optimisation in Schrödinger LiveDesign; Accuracy of toxicity predictions in Exscientia and BenevolentAI should be improved
Additional comments and suggestions	Integration with LIMS would be beneficial; Incorporating customisable reporting tools could help	Better training and support resources are needed, as more guidance for less familiar users of AI and ML	Integration with other bioinformatics tools would enhance utility; Seamless integration for genetic data analysis is recommended	More robust data visualisation capabilities are needed; better visualisation helps interpret complex results	Enhancing collaboration features could benefit larger research teams and facilitate seamless work within the same environment

As can be drawn from Table 4, the experts' feedback suggests that while applications like Insilico Pharma.AI and Atomwise are strong performers with user-friendly interfaces and reliable predictions, others like Schrödinger LiveDesign, Exscientia, and BenevolentAI need improvements in user accessibility, computational efficiency, and accuracy in certain areas. Cyclica, though praised for its simplicity, may be better suited for less complex tasks. Enhancements in integration, visualisation, and support resources are recommended to further advance the utility and adoption of these AI-driven no-code applications in drug discovery.

5. Discussion

The findings of this study underscore the varying effectiveness and usability of selected AI-driven no-code applications in drug discovery, providing valuable insights into their current capabilities and potential areas for improvement. The analysis revealed that Insilico Medicine's Pharma.AI and Atomwise generally outperformed the other applications across multiple metrics, including usability, reliability, and accuracy, suggesting their broad applicability in various stages of drug discovery. Schrödinger's LiveDesign, while resource-intensive, demonstrated a high level of accuracy, making it suitable for environments where precision is critical. In contrast, BenevolentAI and Exscientia showed limitations in both usability and predictive accuracy, indicating a need for further development to enhance their practical utility. Cyclica, noted for its user-friendliness, appears more suited for simpler tasks due to its lower scalability and resource utilisation capabilities.

The study's findings align with existing literature that highlights the potential of AI-driven platforms to revolutionise drug discovery processes. Previous research has shown that AI can significantly improve the efficiency and accuracy of target identification and virtual screening [10, 11]. The strong performance of Insilico Pharma.AI and Atomwise in these areas corroborates these findings, demonstrating their capacity to provide reliable predictions that align with experimental data [30]. These applications leverage advanced machine learning models, such as deep learning and support vector machines, to analyse vast datasets and predict drug-target interactions accurately [10, 16]. However, the observed limitations of BenevolentAI and Exscientia in predictive accuracy, particularly in toxicity predictions, highlight ongoing challenges in the field. As noted by [14], the accuracy of AI predictions is heavily dependent on the quality and diversity of the training data. The inconsistent performance of these applications suggests that they may benefit from incorporating more comprehensive datasets and advanced modelling techniques to improve their predictive power. Moreover, the high resource demands of Schrödinger LiveDesign echo concerns raised by [18] regarding the scalability of AI applications in drug discovery, which can hinder their broader adoption in resource-constrained environments.

Practical relevance. The practical relevance of these findings lies in their potential to guide pharmaceutical companies and research institutions in selecting the most suitable AI-driven no-code applications for their drug discovery projects. By identifying the strengths and weaknesses of each application, stakeholders can make informed decisions about which tools to integrate into their workflows, thereby optimising efficiency and resource allocation. Applications like Insilico Pharma.AI and Atomwise, with their demonstrated reliability and usability, can serve as valuable assets in accelerating the drug discovery process, reducing time-to-market for new therapeutic drugs, and ultimately improving patient outcomes.

Research limitations. The evaluation's scope was limited to a few applications and scenarios, which may not accurately represent the full potential of AI-driven no-code applications in drug discovery. The reliance on expert interviews introduces subjective bias, and the tested scenarios may not capture the complexities of real-world drug discovery projects, thereby limiting the results' generalisability.

Prospects for further research. Future research should aim to expand the scope of evaluation by including a broader range of AI-driven applications and testing them in more diverse and complex drug discovery scenarios. Investigating the integration of these applications with other advanced technologies, such as quantum computing and big data analytics, could provide deeper insights into their potential and limitations. Longitudinal studies assessing the long-term impact of these tools on drug discovery timelines, costs, and success rates would also be valuable. Furthermore, exploring user-centred design improvements based on feedback from a wider range of stakeholders, including scientists, clinicians, and data analysts, could enhance the usability and adoption of these tools.

6. Conclusion

The results indicated that Insilico Medicine's Pharma.AI and Atomwise are robust performers across various metrics, demonstrating strong usability and reliable predictive accuracy in tasks such as target identification and lead optimisation. Schrödinger's LiveDesign, while accurate, requires significant computational resources, which may limit its applicability in resource-constrained environments. In contrast, BenevolentAI and Exscientia were found to have limitations in both usability and predictive accuracy, particularly in toxicity prediction, highlighting areas for future improvement. Cyclica, noted for its ease of use, appears best suited for simpler, less resource-intensive tasks due to its lower scalability and resource efficiency.

These findings underscore the importance of selecting the right AI-driven tools tailored to specific drug discovery tasks and organisational needs. As AI continues to advance, integrating these applications with existing bioinformatics tools and enhancing their capabilities through better data quality, efficient resource utilisation, and user-centred design will be crucial for maximising their potential. Addressing these challenges will not only improve the functionality of these tools but also accelerate the drug discovery process, reduce costs, and enhance the overall success rate of developing new therapeutic drugs. Future research should focus on expanding the scope of evaluation to include a wider range of appli-

cations, scenarios, and real-world case studies, as well as exploring innovative ways to improve the accuracy, scalability, and usability of these tools.

Conflict of interest

The authors of this research article explicitly declare that they do not have any actual or potential conflicts of interest that could compromise the validity and reliability of the research findings presented in this article.

Funding

The study was performed without financial support.

Data availability

The manuscript does not contain any accompanying data.

Use of artificial intelligence

The authors explicitly affirm that they refrained from employing artificial intelligence technologies in the creation of this manuscript on the implementation of AI and machine learning in drug discovery and development. However, AI-driven applications were utilised in the research to evaluate their functionality and effectiveness in various drug discovery tasks.

References

1. Austin, D., Hayford, T. (2021). Research and development in the pharmaceutical industry. CBO. Available at: https://www.cbo.gov/publication/57126

2. Schuhmacher, A., Hinder, M., von Stegmann und Stein, A., Hartl, D., Gassmann, O. (2023). Analysis of pharma R&D productivity – a new perspective needed. Drug Discovery Today, 28 (10), 103726. https://doi.org/10.1016/j.drudis.2023.103726

3. Blanco-González, A., Cabezón, A., Seco-González, A., Conde-Torres, D., Antelo-Riveiro, P., Piñeiro, Á., Garcia-Fandino, R. (2023). The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. Pharmaceuticals, 16 (6), 891. https://doi.org/10.3390/ph16060891

4. Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. Drug Discovery Today, 26 (1), 80–93. https://doi.org/10.1016/j.drudis.2020.10.010

5. Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S. et al. (2021). Artificial intelligence: A powerful paradigm for scientific research. The Innovation, 2 (4), 100179. https://doi.org/10.1016/j.xinn.2021.100179

6. Singh, S., Kumar, R., Payra, S., Singh, S. K. (2023). Artificial Intelligence and Machine Learning in Pharmacological Research: Bridging the Gap Between Data and Drug Discovery. Cureus. https://doi.org/10.7759/cureus.44359

7. Kim, H., Kim, E., Lee, I., Bae, B., Park, M., Nam, H. (2020). Artificial Intelligence in Drug Discovery: A Comprehensive Review of Data-driven and Machine Learning Approaches. Biotechnology and Bioprocess Engineering, 25 (6), 895–930. https://doi.org/10.1007/s12257-020-0049-y

8. Ghislat, G., Hernandez-Hernandez, S., Piyawajanusorn, C., Ballester, P. J. (2024). Data-centric challenges with the application and adoption of artificial intelligence for drug discovery. Expert Opinion on Drug Discovery, 19(11), 1297–1307. https://doi.org/10.1080/17460441.2024.2403639

9. Kolluri, S., Lin, J., Liu, R., Zhang, Y., Zhang, W. (2022). Machine Learning and Artificial Intelligence in Pharmaceutical Research and Development: a Review. The AAPS Journal, 24 (1). https://doi.org/10.1208/s12248-021-00644-3

10. Lavecchia, A. (2015). Machine-learning approaches in drug discovery: methods and applications. Drug Discovery Today, 20 (3), 318–331. https://doi.org/10.1016/j.drudis.2014.10.012

11. Gomes, J., Ramsundar, B., Feinberg, E. N., Pande, V. S. (2017). Atomic convolutional networks for predicting protein-ligand binding affinity. https://doi.org/10.48550/arXiv.1703.10603

12. Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V. et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature Biotechnology, 37 (9), 1038–1040. https://doi.org/10.1038/s41587-019-0224-x

13. Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M. et al. (2020). A deep learning approach to antibiotic discovery. Cell, 181 (2), 475–483. https://doi.org/10.1016/j.cell.2020.01.021

14. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G. et al. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18 (6), 463–477. https://doi.org/10.1038/s41573-019-0024-5

15. Libbrecht, M. W., Noble, W. S. (2015). Machine learning applications in genetics and genomics. Nature Reviews Genetics, 16 (6), 321–332. https://doi.org/10.1038/nrg3920

16. Zhou, J., Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning–based sequence model. Nature Methods, 12 (10), 931–934. https://doi.org/10.1038/nmeth.3547

17. Tiwary, S., Levy, R., Gutenbrunner, P., Salinas Soto, F., Palaniappan, K. K., Deming, L. et al. (2019). High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. Nature Methods, 16 (6), 519–525. https://doi.org/10.1038/s41592-019-0427-6

18. Casadio, R., Martelli, P. L., Savojardo, C. (2022). Machine learning solutions for predicting protein–protein interactions. WIREs Computational Molecular Science, 12(6). https://doi.org/10.1002/wcms.1618

19. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005

20. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542 (7639), 115–118. https://doi.org/10.1038/nature21056

21. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T. (2018). The rise of deep learning in drug discovery. Drug Discovery Today, 23 (6), 1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039

22. Wallach, I., Dzamba, M., Heifets, A. (2015). AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. arXiv preprint arXiv:1510.02855. https://doi.org/10.48550/arXiv.1510.02855

23. Xu, Y., Dai, Z., Chen, F., Gao, S., Pei, J., Lai, L. (2015). Deep Learning for Drug-Induced Liver Injury. Journal of Chemical Information and Modeling, 55 (10), 2085–2093. https://doi.org/10.1021/acs.jcim.5b00238

24. Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S. (2017). DeepTox: Toxicity prediction using deep learning. Toxicology Letters, 280, S69. https://doi.org/10.1016/j.toxlet.2017.07.175

25. Cavasotto, C. N., Scardino, V. (2022). Machine Learning Toxicity Prediction: Latest Advances by Toxicity End Point. ACS Omega, 7(51), 47536–47546. https://doi.org/10.1021/acsomega.2c05693

26. Nam, S., Kim, D., Jung, W., Zhu, Y. (2022). Understanding the Research Landscape of Deep Learning in Biomedical Science: Scientometric Analysis. Journal of Medical Internet Research, 24 (4), e28114. https://doi.org/10.2196/28114

27. Kavalci, E., Hartshorn, A. (2023). Improving clinical trial design using interpretable machine learning based prediction of early trial termination. Scientific Reports, 13 (1). https://doi.org/10.1038/s41598-023-27416-7

28. Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., Socher, R. (2021). Deep learning-enabled medical computer vision. Npj Digital Medicine, 4 (1). https://doi.org/10.1038/s41746-020-00376-2

29. Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., Chen, I. Y., Ranganath, R. (2020). A Review of Challenges and Opportunities in Machine Learning for Health. AMIA Joint Summits on Translational Science proceedings, 191–200. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233077/pdf/3268580.pdf

30. Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., Chen, I. Y., Ranganath, R. (2019). Practical guidance on artificial intelligence for health-care data. The Lancet Digital Health, 1 (4), e157–e159. https://doi.org/10.1016/s2589-7500(19)30084-6

31. Ghosh, S., Das, N., Chakraborty, S., Chakrabarti, A. (2020). Understanding deep learning techniques for image segmentation. arXivLabs. https://doi.org/10.48550/arXiv.1907.06119

32. Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366 (6464), 447–453. https://doi.org/10.1126/science.aax2342

33. Jiménez-Luna, J., Grisoni, F., Weskamp, N., Schneider, G. (2021). Artificial intelligence in drug discovery: recent advances and future perspectives. Expert Opinion on Drug Discovery, 16 (9), 949–959. https://doi.org/10.1080/17460441.2021.1909567

34. Lundberg, S. M., Lee, S. I. (2017). A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems, 4768–4777. https://doi.org/10.48550/arXiv.1705.07874

35. Ribeiro, M. T., Singh, S., Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

36. Bauer, C. (2020). Multi-Method Evaluation: Leveraging Multiple Methods to Answer What You Were Looking For. Proceedings of the 2020 Conference on Human Information Interaction and Retrieval (CHIIR '20). Association for Computing Machinery, 472–474. https://doi.org/10.1145/3343413.3378015

37. Pharma.AI. Insilico Medicine's. Available at: https://pharma.ai/

38. Atomwise. Available at: https://www.atomwise.com/

39. Schrödinger's. LiveDesign. Available at: https://www.schrodinger.com/

40. Exscientia. Available at: https://www.exscientia.com/

41. Benevolent AI. Available at: https://www.benevolent.com//

42. Cyclica. Available at: https://cyclicarx.com/?lang=en

43. The jamovi project (2024). jamovi (Version 2.5) [Computer Software]. Available at: https://www.jamovi.org/

Received 18.09.2024
Received in revised form 04.12.2024
Accepted 19.12.2024
Published 30.12.2024

Iryna Nizhenkovska, Doctor of Medical Sciences, Professor, Department of Medicinal Chemistry and Toxicology, Bogomolets National Medical University, T. Shevchenko blvd., 13, Kyiv, Ukraine, 01601

Tetyana Reva, Doctor of Pedagogical Sciences, Professor, Department of Analytical, Physical and Colloid Chemistry, Bogomolets National Medical University, T. Shevchenko blvd., 13, Kyiv, Ukraine, 01601

Olena Kuznetsova*, PhD, Associate Professor, Department of Medicinal Chemistry and Toxicology, Bogomolets National Medical University, T. Shevchenko blvd., 13, Kyiv, Ukraine, 01601

Oleksii Nizhenkovskyi, PhD, Department of Pharmacy and Industrial Technology of Drugs, Bogomolets National Medical University, T. Shevchenko blvd., 13, Kyiv, Ukraine, 01601

Oksana Chkhalo, PhD, Associate Professor, Department of Analytical, Physical and Colloid Chemistry, Bogomolets National Medical University, T. Shevchenko blvd., 13, Kyiv, Ukraine, 01601

***Corresponding author:** Olena Kuznetsova, e-mail: olena.kuznetsova@nmu.ua*

Appendix 1

Researcher Designed Scenarios for Five Standardised Tests to Evaluate AI-Driven No-Code Applications in Drug Discovery

Scenario 1: Target Identification

(Used to test the functionality of apps such as Insilico Medicine's Pharma.AI, Exscientia & BenevolentAI)

Objective: Evaluate the application's ability to identify potential drug targets based on disease-related proteins.

Description:

Disease Selection:

• Alzheimer's Disease: A neurodegenerative disorder characterised by the accumulation of amyloid-beta plaques and tau tangles in the brain.

• Breast Cancer: A common cancer type where mutations in certain genes, such as BRCA1 and HER2, play a critical role in the disease's progression.

Dataset Preparation:

Genomic and Proteomic Data:

For Alzheimer's Disease:

• Genomic Data: Whole-genome sequencing data from Alzheimer's patients, including known mutations in genes such as APP, PSEN1, and APOE.

• Proteomic Data: Protein expression profiles from Alzheimer's brain tissues, focusing on proteins like amyloid-beta (Aβ) and tau.

For Breast Cancer:

• Genomic Data: Whole-exome sequencing data from breast cancer patients, including mutations in *BRCA1*, *BRCA2*, and *TP53*.

• Proteomic Data: Protein expression profiles from tumor samples, with an emphasis on proteins like HER2 (ERBB2) and estrogen receptor (ER).

Task:

The application is to be tasked with analysing these genomic and proteomic datasets to identify novel drug targets:

• For Alzheimer's Disease: The application might be supposed to identify proteins involved in the amyloid-beta pathway or tau phosphorylation as potential drug targets.

• For Breast Cancer: The application could be expected to pinpoint mutations in PIK3CA or overexpression of HER2 as potential targets for therapeutic intervention.

Validation:

• Known Targets: The application's identified targets is to be compared against known targets, such as those listed in databases like DrugBank, OMIM, or the Therapeutic Target Database (TTD).

• Novel Targets: For novel targets, additional validation is to be performed using literature searches and experimental data to determine their relevance and potential for drug development.

Evaluation Metrics:

Accuracy: The proportion of correctly identified targets relative to known targets.

Novelty: The application's ability to identify previously unknown targets.

Speed: The time taken to process the datasets and generate results.

Scenario 2: Virtual Screening

(Used to test the functionality of apps such as Insilico Medicine's Pharma.AI & Atomwise)

Objective: Assess the application's capability to screen large chemical libraries for compounds with potential binding affinity to a target protein.

Description:

Target Protein Selection:

• EGFR (Epidermal Growth Factor Receptor): A well-known target in non-small cell lung cancer (NSCLC). Mutations in EGFR can lead to uncontrolled cell proliferation, making it a prime target for cancer therapies.

• BACE1 (Beta-Secretase 1): A target for Alzheimer's disease, involved in the production of amyloid-beta peptides, which aggregate to form plaques in the brain.

• SARS-CoV-2 Main Protease (Mpro): A critical enzyme in the life cycle of the SARS-CoV-2 virus, making it a target for COVID-19 antiviral drug development.

Chemical Library Selection:

• ZINC15 Database: A publicly available chemical library containing over 230 million purchasable compounds. For this scenario, a subset of 1,000,000 diverse compounds will be used for screening.

• ChEMBL: A database of bioactive drug-like small molecules with over 2 million compounds, widely used in drug discovery research.

• Enamine REAL Database: A chemical library containing over 3 billion synthetically accessible compounds, suitable for large-scale virtual screening projects.

Task:

• The application is to be tasked with virtually screening a library of 1,000,000 chemical compounds against a specific target protein.

• The virtual screening process involves docking the compounds into the binding site of the target protein and predicting their binding affinities.

• The output will be a ranked list of compounds, with those predicted to have the highest binding affinity at the top.

Validation:

• The top-ranked compounds are to be cross-referenced with known inhibitors or binders of the target protein from databases such as DrugBank, PubChem, and BindingDB.

• Experimentally validated binding affinities are to be compared to the predicted affinities to assess the accuracy of the virtual screening.

Evaluation Metrics:

Accuracy: The percentage of top-ranked compounds that exhibit experimentally validated binding affinity.

Efficiency: The computational resources and time required to complete the virtual screening.

Scalability: The application’s ability to handle larger libraries without significant performance degradation.

Appendix 2

Expert Interview Questionnaire

1. What were your initial impressions of the overall functionality and usability of each AI-driven no-code application you tested?

2. Do you believe that the applications tested are capable of offering reliable and accurate results for drug discovery professionals? Please provide specific examples or scenarios to support your reasoning.

3. What are the main areas where these applications could be improved to better support drug discovery efforts, particularly in terms of performance metrics like speed, accuracy, and resource utilisation?

4. Do you have any additional comments or suggestions regarding the use and potential integration of AI-driven no-code applications in drug discovery that were not covered in the previous questions?

Scenario 3: Drug Design

(Used to test the functionality of apps such as Insilico Medicine's Pharma.AI, Atomwise, Schrödinger's LiveDesign, Exscientia, BenevolentAI, Cyclica)

Objective: Evaluate the application's effectiveness in designing new drug molecules optimised for specific properties such as binding affinity, solubility, bioavailability, and selectivity.

Description:

Target Protein Selection:

• HIV-1 Protease: A crucial enzyme in the life cycle of the HIV virus, making it a prime target for antiretroviral drugs.

• PD-L1 (Programmed Death-Ligand 1): A target in immuno-oncology therapies, where inhibiting the PD-1/PD-L1 interaction can enhance the immune response against tumors.

• ACE2 (Angiotensin-Converting Enzyme 2): A receptor for SARS-CoV-2, making it a target for designing antiviral drugs against COVID-19.

Design Objectives:

• *Binding Affinity:* The application should design molecules with strong binding affinity to the selected target protein.

• *Solubility:* The designed molecules should have high solubility in aqueous environments to ensure adequate bioavailability.

• *Bioavailability:* The molecules should be optimised for oral bioavailability, considering factors such as molecular weight, lipophilicity, and hydrogen bonding potential.

• *Selectivity:* The application should design molecules that are selective for the target protein, minimising off-target interactions that could lead to side effects.

Assessment Criteria:

Lipinski's Rule of Five: The designed molecules are to be assessed for drug-likeness based on Lipinski's Rule of Five, which includes criteria like:

• No more than 5 hydrogen bond donors (OH and NH groups).

• No more than 10 hydrogen bond acceptors (N and O atoms).

• A molecular weight under 500 Da.

• A partition coefficient (LogP) less than 5.

Additional Optimisation Goals:

Pharmacokinetics: The application should also consider pharmacokinetic properties such as half-life and metabolism.

Toxicity: The molecules should be designed to avoid known toxicophores and minimise predicted toxicity.

Evaluation Metrics:

Design Quality:

Definition: The degree to which the designed molecules meet the desired properties (binding affinity, solubility, bioavailability, selectivity).

Creativity:

Definition: The novelty of the molecular structures proposed by the application.

Optimisation Efficiency:

Definition: The time and computational resources required to generate optimised drug candidates.

Scenario 4: Lead Optimisation

(Used to test the functionality of apps such as Insilico Medicine's Pharma.AI, Atomwise, Schrödinger's LiveDesign, Exscientia, BenevolentAI, Cyclica)

Objective: Test the application's ability to refine lead compounds to enhance their efficacy, selectivity, and safety.

Description:

Selection of Target Protein and Lead Compounds:

Target Protein: VEGFR-2 (Vascular Endothelial Growth Factor Receptor-2)

Involved in angiogenesis, making it a target for cancer therapies, particularly in inhibiting tumor blood supply.

Lead Compound Example: Sunitinib

A multi-targeted receptor tyrosine kinase (RTK) inhibitor that has moderate activity against VEGFR-2 but requires optimisation to improve selectivity and reduce cardiotoxicity.

Target Protein: BACE1 (Beta-Secretase 1)

Involved in the production of amyloid-beta, making it a target for Alzheimer's disease.

Lead Compound Example: LY2886721

A BACE1 inhibitor that showed promise in early trials but was discontinued due to liver toxicity.

Target Protein: MMP-9 (Matrix Metalloproteinase-9)

Involved in tissue remodeling and implicated in cancer metastasis and chronic inflammatory diseases.

Lead Compound Example: Marimastat

A broad-spectrum MMP inhibitor that was discontinued due to musculoskeletal toxicity, requiring selective optimisation.

Optimisation Objectives:

Potency Improvement:

Increase the binding affinity of the lead compounds to their respective target proteins. For example, optimising Sunitinib to achieve stronger inhibition of VEGFR-2.

Selectivity Enhancement:

Reduce off-target interactions that could lead to side effects. For instance, modifying the structure of LY2886721 to reduce its interaction with liver enzymes, thus lowering hepatotoxicity.

Safety Improvement:

Optimise pharmacokinetic properties, including absorption, distribution, metabolism, and excretion (ADME), to enhance the safety profile. For Marimastat, structural changes could be explored to reduce its impact on musculoskeletal tissues.

Lead Compound Optimisation Process:

Structure-Based Drug Design:

The application will use the 3D structures of the lead compounds bound to their target proteins to identify regions of the molecule that can be modified to improve binding affinity and selectivity.

Computational Screening:

The application will generate and virtually screen multiple analogues of the lead compound, predicting their binding affinities, selectivity profiles, and ADME characteristics.

Iterative Optimisation:

The application will iteratively refine the lead compounds, using feedback from each round of screening to enhance efficacy and reduce potential toxicity.

Evaluation Metrics:

Improvement in Potency:

Definition: The increase in binding affinity or activity compared to the original lead compounds.

Selectivity:

Definition: The ability to reduce off-target interactions while maintaining or improving efficacy.

Safety Profiles:

Definition: Predicted improvements in pharmacokinetic properties and reduced toxicity.

Scenario 5: Toxicity Prediction

(Used to test the functionality of apps such as Insilico Medicine's Pharma.AI, Schrödinger's LiveDesign, Exscientia, Cyclica)

Objective: Assess the application's capability to predict potential toxic effects of drug candidates, focusing on common toxicity endpoints such as hepatotoxicity, cardiotoxicity, and genotoxicity.

Description:

Selection of Drug Candidates:

Candidate 1: Acetaminophen (Paracetamol) Derivatives

Potential Toxicity: Hepatotoxicity

Background: Acetaminophen is widely used as an analgesic and antipyretic, but overdoses can cause severe liver damage. The application will be tasked with predicting hepatotoxicity in novel acetaminophen derivatives.

Candidate 2: Doxorubicin Analogues

Potential Toxicity: Cardiotoxicity

Background: Doxorubicin, an anthracycline antibiotic used in chemotherapy, is known for its effectiveness but also for its dose-dependent cardiotoxicity. The application will predict cardiotoxic effects in new analogues.

Candidate 3: Nitrosourea-Based Compounds

Potential Toxicity: Genotoxicity

Background: Nitrosoureas are alkylating agents used in cancer treatment, but they can cause DNA damage leading to mutagenesis and genotoxicity. The application will assess the genotoxic potential of new nitrosourea derivatives.

Toxicity Endpoints:

• *Hepatotoxicity:* Predict the likelihood of liver enzyme elevation (e.g., ALT, AST) and liver damage, which could lead to conditions like drug-induced liver injury (DILI).

• *Cardiotoxicity:* Predict the potential for QT interval prolongation, cardiomyopathy, or arrhythmias, which could lead to heart failure or sudden cardiac death.

• *Genotoxicity:* Predict the potential for DNA damage, mutagenesis, and chromosomal aberrations, which could result in carcinogenicity or teratogenic effects.

Toxicity Prediction Process:

Data Input:

The application is to be provided with the chemical structures and relevant physicochemical properties of the drug candidates.

Prediction Models:

The application will use in silico models to predict toxicological endpoints based on known toxicity data, structure-activity relationships (SAR), and machine learning algorithms trained on large toxicity datasets (e.g., Tox21, ToxCast).

Output:

The application will generate a toxicity profile for each drug candidate, indicating the likelihood of various toxic effects, such as hepatotoxicity, cardiotoxicity, and genotoxicity.

Evaluation Metrics:

Accuracy:

Definition: The percentage of correctly predicted toxic and non-toxic compounds compared to experimental data.

Definition:

Sensitivity: The ability to correctly identify compounds that are toxic.

Specificity: The ability to correctly identify compounds that are non-toxic.

Predictive Power:

Definition: The application's ability to predict specific types of toxicity with high confidence.