Review and Progress Open Access

AI Based Drug Screening Process: From Data Mining to Candidate Drug Validation

Wang wei

Institute of Life Science,Jiyang College of Zhejiang A&F University, zhuji, 311800, China

Corresponding author email: 2741098603@qq.com

Bioscience Method, 2024, Vol.15, No.1 doi: 10.5376/bm.2024.15.0005

Received: 11 Jan., 2024

Accepted: 16 Feb., 2024

Published: 28 Feb., 2024

Preferred citation for this article:

Wang W., 2024, AI based drug screening process: from data mining to candidate drug validation, Bioscience Method, 14(1): 37-49 (doi: 10.5376/bm.2024.15.0005)

Abstract With the rapid development of artificial intelligence (AI) technology, its application in drug research and development is becoming increasingly widespread. This study introduces the advantages of AI technology in drug screening, such as fast processing and analysis of large amounts of data, improving screening accuracy, and reducing research and development costs. Discussed the shortcomings in the current AI drug screening process, such as data dependence, insufficient model interpretability, and legal and ethical issues. Intended to explore the AI based drug screening process, from data mining to candidate drug validation. I hope to provide a comprehensive and systematic perspective for researchers and practitioners in the field of drug development by deeply understanding the advantages, disadvantages, and challenges faced by AI technology in drug screening, and proposing corresponding solutions, in order to guide them to better utilize AI technology to accelerate the drug development process.

Keywords Artificial intelligence; Drug screening; Data mining; Candidate drug validation; Drug development

Drug development is a crucial link in the pharmaceutical field, with immeasurable value in improving human health and treating diseases (Wu et al., 2019). The discovery and development of new drugs can provide new treatment plans for various diseases, improve the treatment effect of diseases, reduce treatment costs, and even find breakthroughs for some currently incurable diseases. Drug development can not only improve the quality of life of individual patients, but also have a positive impact on the overall health level of society. Ping has a positive impact.

However, the traditional drug development process is time-consuming and labor-intensive, and the success rate is often not satisfactory. This is mainly attributed to the complexity of biological systems, the complexity of drug target interactions, and the vast space for candidate drugs. With the rapid development of biotechnology, the standards and requirements for drug research and development are also constantly improving, which brings greater challenges to drug research and development. Therefore, developing more efficient and accurate drug screening methods has become an urgent task.

In recent years, the rapid development of artificial intelligence (AI) technology has brought revolutionary changes to the field of drug research and development (Mak and Pichika, 2019). AI can use technologies such as deep learning and machine learning to extract valuable information from massive data, predict drug target interactions, evaluate drug efficacy and safety. This not only greatly improves the efficiency and accuracy of drug screening, but also provides new ideas and directions for drug development.

The rise of AI technology, especially the development of data mining and machine learning technologies, has provided new possibilities and enormous potential for drug screening. Data mining technology can efficiently process and analyze a large amount of biomedical data, extract information related to drug activity, safety, etc. Machine learning algorithms can utilize this data to construct accurate prediction models, helping researchers quickly screen potential candidate drugs.

This study aims to explore how to combine data mining and machine learning techniques to construct an efficient AI drug screening process. Through in-depth analysis and discussion of the advantages, challenges, and future development trends of this process, it is hoped to provide useful reference and inspiration for researchers in the field of drug research and development, and promote the further development and application of AI based drug screening technology.

1 Overview of Drug Screening Process

1.1 Basic process and key steps of drug screening

The drug screening process is the core link in drug development, which involves accurately selecting compounds with therapeutic potential from a massive selection of candidate drugs. This process typically begins with in-depth research on specific diseases or symptoms to clarify their biological mechanisms and potential drug targets (Figure 1). Researchers will screen potential active candidate drugs from a wide range of drug or compound libraries based on these targets. This preliminary screening process is usually completed through in vitro experiments such as high-throughput screening techniques and cell models to quickly evaluate the affinity and activity of candidate drugs with target molecules.

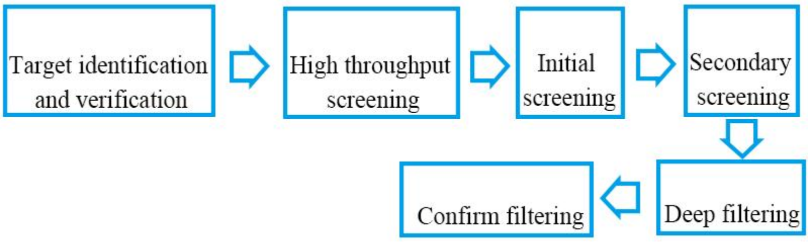

Figure 1 Steps of drug screening

Pan et al. (2019) developed a novel high-throughput screening method using gas chromatography-high-resolution mass spectrometry (GC-HRMS) technology for the screening of 288 drugs and toxins in human blood. This method allows for rapid detection and identification of many forensic important drugs and toxins, such as abused drugs (such as cocaine, amphetamines, synthetic cannabinoids, opioids, hallucinogens), sedatives and hypnotics, antidepressants, nonsteroidal anti-inflammatory drugs, insecticides (such as acaricides, fungicides, insecticides, nematicides), and cardiovascular drugs.

Costa et al. (2019) developed a two-step drug screening process, including rapid screening by paper spray method, and then confirmed by liquid chromatography/mass spectrometry (LC/MS). This method demonstrates the potential application of testing drug compliance from fingerprints. This method first uses paper spray analysis to quickly screen a large number of samples, and then uses LC/MS to confirm any controversial results, especially the screening and confirmation of the antipsychotic drug quetopine, demonstrating the practicality of this method.

Lin and Zhou (2022) proposed a drug candidate screening scheme based on machine learning methods to improve the efficiency of drug screening. This method can not only discover appropriate compounds, but also reveal the potential impact of molecular descriptors (i.e. feature values) on the properties of compounds. This work involves training an accurate prediction model based on independent variables (i.e. eigenvalues) and dependent variables (i.e. biological activity values or ADMET attributes), then using feature interpretation algorithms to select features that have a significant impact on dependent variables, finally finding approximate optimal values for these important features, and analyzing numerical ranges that are beneficial for obtaining better biological activity and ADMET attributes.

However, relying solely on in vitro experiments is not sufficient. The in-depth screening stage requires researchers to further explore the activity, pharmacokinetic characteristics, and safety of candidate drugs in vivo. This stage of research is usually more complex and time-consuming, and requires the use of animal models or clinical trials to verify the actual effects of candidate drugs.

During the entire screening process, selecting appropriate screening models, effectively analyzing and interpreting experimental data, and continuously optimizing the structure and properties of candidate drugs are all crucial steps. To ensure the accuracy and reliability of the screening results, researchers often need to conduct multiple rounds of screening and validation, including repeated experiments under different experimental conditions and cross validation using different models.

1.2 The role and application of AI in drug screening

AI (Artificial Intelligence) is playing an increasingly important role in drug screening. Its application not only improves screening efficiency, but also reduces research and development costs, bringing revolutionary changes to the field of drug development (Mohanty et al., 2020). AI can automate the processing and analysis of large-scale biomedical data, including genomics, proteomics, drug chemical structures, etc., in order to quickly identify potential drug targets related to specific diseases. This data mining and integration capability greatly surpasses traditional manual methods, providing researchers with more comprehensive and in-depth information.

AI performs excellently in predicting drug target interactions. Hessler and Baringhaus (2018) explored the key role of AI in drug design, particularly how artificial neural networks such as deep neural networks drive various aspects of drug discovery, in their study. The application of AI in quantitative structure-activity relationship (QSAR) has demonstrated its strength in predicting physicochemical and ADMET (absorption, distribution, metabolism, excretion, and toxicology) properties. In addition, AI has demonstrated its strong ability to generate new bioactive molecules with desired properties, laying the foundation for its strength in drug discovery.

Walters and Barzilay (2021) critically evaluated the application of AI in drug discovery in their review, exploring its application in analyzing high-content screening data, designing and synthesizing new molecules, and other drug discovery areas. They discussed the different fields in which AI is applied in drug discovery, including attribute prediction, molecular generation, image analysis, and organic synthesis planning.

Deng et al. (2021) provided a review on the application and technology of AI in drug discovery. This study first provides an overview of drug discovery and its related applications, and then discusses common data resources, molecular representations, and benchmark platforms. AI technology is divided into model architectures and learning paradigms. They review the application of AI in drug discovery and provide a GitHub repository containing relevant papers (and applicable code) as a learning resource.

Therefore, AI has the potential in drug screening and discovery, especially in molecular property prediction, molecular generation, and applications combined with synthesis planning and drug design. AI can also be used for virtual screening and experimental design, and even plays an important role in personalized treatment.

1.3 The importance of data mining and machine learning in drug development

The importance of data mining and machine learning in drug development is self-evident. They provide a new perspective and tools for drug development, driving rapid development in the field of drug development. Data mining technology can extract information related to drug development from massive biomedical data. These data may come from multiple levels such as genomics, transcriptomics, proteomics, metabolomics, etc., covering rich content such as disease pathogenesis, drug action mechanisms, and clinical information of patients (Kavakiotis et al., 2017). Through data mining, researchers can gain a deeper understanding of the nature of diseases and the ways drugs work, providing strong data support for drug development.

Machine learning algorithms can intelligently analyze and predict these data. Based on a large amount of training data, machine learning models can learn the complex relationship between drugs and diseases, predict key attributes such as the effectiveness and safety of candidate drugs. This predictive ability not only greatly improves the efficiency and accuracy of drug screening, but also helps to reduce the failure rate and risk in clinical trials.

Data mining and machine learning can also provide personalized solutions for drug development. By conducting in-depth analysis of patient genomics, clinical data, etc., researchers can develop personalized treatment plans for each patient, improving treatment effectiveness and quality of life. The concept of precision medicine is gradually changing traditional medical models, bringing patients a better treatment experience.

2 Data Mining and Feature Selection

2.1 Data types and sources required for drug screening

Drug screening is a complex process that requires support from multiple data types. To ensure the accuracy and effectiveness of screening, researchers need to collect and analyze data from multiple sources and types. From the perspective of data types, genomic and transcriptomic data provide researchers with information about genes and gene expression, which is crucial for identifying genes and gene pathways associated with specific diseases. Meanwhile, proteomic data involves protein expression, structure, and function, which is crucial for the discovery of drug targets. Metabolomics data describes the metabolic processes and metabolites within organisms, providing valuable clues for understanding the mechanisms of disease occurrence and development. In addition, clinical data is an important basis for evaluating the efficacy and safety of drugs, including key information such as the patient's medical history, symptoms, and treatment effectiveness. Drug chemistry and biological activity data provide information about drug structure, mechanism of action, and biological activity, which is the basis for drug screening (Figure 2).

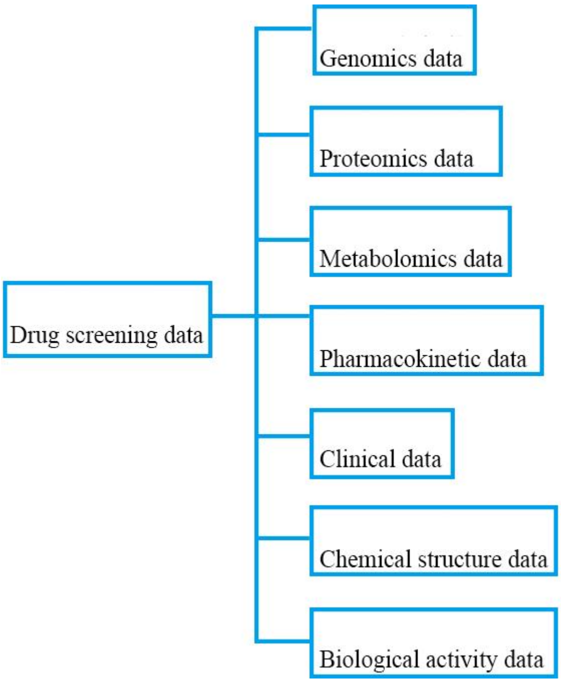

Figure 2 Types of drug screening data

In terms of data sources, public databases are an important way for researchers to obtain biomedical data. Dogan (2018) found that NCBI databases such as GeneBank, UniProt, and MetaboLights store a large amount of biomedical data for researchers to access free of charge or for a fee. In addition, Burton et al. (2017) believe that research institutions are also an important source of data. Research institutions and teams around the world have accumulated a large amount of experimental data and research results in drug development. By collaborating or purchasing data with these institutions, researchers can obtain valuable data resources. Walke et al. (2023) argue that clinical trials are a crucial step in evaluating drug efficacy and safety, and the data generated is crucial for drug screening and development. Long term registration and follow-up of patients can also collect valuable data on disease progression and treatment effectiveness, providing strong support for drug screening.

2.2 Application of data mining technology in drug development

In the drug discovery stage, data mining techniques are widely applied in data analysis in fields such as genomics, proteomics, and metabolomics. By deeply mining these large-scale biomedical data (Yang et al., 2020), researchers can identify genes, proteins, or metabolites associated with specific diseases, thereby identifying potential drug targets. This greatly accelerates the speed of drug discovery and improves the success rate of research and development.

In the drug design phase, data mining techniques can help researchers predict the biological activity, pharmacokinetic characteristics, and possible side effects of candidate drugs. In the clinical trial stage, by analyzing and mining a large amount of clinical data, researchers can evaluate the efficacy and safety of drugs, discover the correlation between drugs and diseases, and predict patient reactions to different drugs. In the application of drug market, data mining technology can help pharmaceutical companies analyze customer needs and concerns, optimize product strategies and marketing plans.

The application of data mining technology in drug development involves multiple aspects, including but not limited to the detection of adverse drug reactions, drug safety monitoring, pharmacodynamics, and prediction of drug interactions. For example, Karimi et al. (2015) reviewed how to use data mining and related computer science technologies from different data sources (including spontaneous reporting databases, electronic health records, and medical literature) to identify signals of adverse drug reactions in the field of drug safety. Wilson et al. (2003) discussed the potential use of data mining and knowledge discovery for detecting adverse drug events (ADEs) in databases and explored the application of data mining in drug surveillance systems. Harpaz et al. (2012) provided an overview of recent methodological innovations and data sources used to support the discovery and analysis of adverse drug events, emphasizing the importance of data mining techniques in improving drug safety monitoring.

2.3 Optimizing feature selection to improve model accuracy and efficiency

In the data mining process of drug development, feature selection is a crucial step that directly affects the accuracy and efficiency of the model. Optimizing feature selection can not only improve the predictive performance of the model, but also simplify the model, reduce computational complexity, and accelerate the drug development process. Feature selection helps to reduce data dimensionality. In drug development, a large amount of biomedical data is usually generated, which may contain many features unrelated to drug activity. By selecting the most important features, redundant and noisy data can be removed, the model can be simplified, computational complexity can be reduced, and the generalization ability of the model can be improved.

Optimizing feature selection can improve the predictive accuracy and interpretability of the model. Selecting the most representative features can make the model more focused on factors closely related to drug activity, thereby improving the predictive accuracy of the model. This is crucial for drug screening and drug design, as it can help researchers quickly identify potential candidate drugs; Selecting features with clear biological significance can make the model easier to understand and interpret. This is crucial for decision-making and communication in the drug development process, as it can help researchers and decision-makers better understand the results and meaning of the model.

To achieve optimization of feature selection, multiple methods and techniques can be employed. For example, statistical feature selection methods can evaluate the importance of features by calculating their correlation or significance with the target variable. Machine learning algorithms (Cai et al., 2018) such as decision trees, random forests, support vector machines, etc. can also be used for feature selection, selecting the best features by training the model and evaluating the impact of features on model performance.

3 Machine Learning Model Construction

3.1 Basic principles of machine learning in drug screening

The basic principle of machine learning in drug screening is to train a model using a large amount of data, enabling the model to automatically learn and recognize features or patterns related to drug activity. These learned features or patterns can be used to predict the biological activity of new compounds (Yang et al., 2019), thereby accelerating the process of drug screening.

Specifically, machine learning algorithms learn a mapping relationship or function from known drug data through continuous iteration and optimization, which can map the characteristics of a compound (such as chemical structure, physical properties, etc.) to its biological activity. This mapping relationship is learned through training samples and their corresponding labels (such as active or inactive) in the data.

In drug screening, machine learning models are often used to preliminarily screen a large number of candidate compounds. By extracting and encoding the features of candidate compounds, machine learning models can predict the biological activity of these compounds, thereby screening compounds with potential activity and providing a candidate list for subsequent experimental verification. It should be noted that the accuracy and reliability of machine learning models highly depend on the quality and quantity of training data. Therefore, when constructing machine learning models, it is necessary to ensure the accuracy and completeness of training data, and collect as many samples as possible to improve the model's generalization ability.

3.2 Common machine learning algorithms and models

The field of machine learning provides various algorithms and models for processing and analyzing large amounts of data, and is applied in various fields such as image processing, natural language processing, predictive modeling, etc. Ray (2019) briefly reviewed the most commonly used and popular machine learning algorithms, emphasizing the advantages and disadvantages of these machine learning algorithms from an application perspective, in order to help make wise decisions about selecting appropriate learning algorithms to meet specific application needs. Raju et al. (2023) compared various popular supervised learning algorithms, such as SVM, decision tree, random forest, KNN, logistic regression, etc., and tested the efficiency of the algorithms on three different datasets in different fields, aiming to compare different algorithms used on different datasets in different fields to understand the best algorithm and overall best algorithm. Mitchell (2014) focuses on certain machine learning methods commonly used in chemical informatics and quantitative structure-activity relationships (QSAR), including artificial neural networks, random forests, support vector machines, k-nearest neighbors, and naive Bayesian classifiers.

In drug screening and drug development, commonly used machine learning algorithms and models include the following:

Linear regression: used to establish linear relationships between variables and predict the values of continuous variables. In drug development, linear regression can be used to predict continuous indicators such as drug efficacy or toxicity.

Logistic Regression: An algorithm used to establish classification models. By mapping the output of a linear regression model to a probability value, logistic regression can achieve binary or multi classification tasks, such as predicting whether a drug has a certain activity.

Decision Tree: An algorithm based on tree structure used to establish classification or regression models. The decision tree recursively divides the dataset into subsets, makes judgments based on eigenvalues, and constructs a tree like structure. In drug development, decision trees can be used to identify key features that affect drug activity.

Random Forest: An ensemble learning algorithm composed of multiple decision trees. Construct multiple decision trees through random sampling and feature selection, and obtain the final prediction results through voting or averaging. Random forests have high accuracy and robustness, making them suitable for processing large-scale datasets (Lei et al., 2021).

Support Vector Machine (SVM): an algorithm used to establish classification and regression models. By mapping data to a high-dimensional space and finding a hyperplane to maximize the spacing between different categories, classification tasks can be achieved. SVM has strong generalization ability and robustness, and can handle high-dimensional data.

Naive Bayes: A classification algorithm based on Bayesian theorem. It assumes that features are independent of each other and classifies them by calculating a posterior probability. Naive Bayes algorithm is simple and efficient, suitable for processing large-scale datasets and high-dimensional data.

K-means Clustering: An unsupervised learning algorithm used to partition a dataset into K non overlapping clusters. By iteratively calculating the distance between each sample and the cluster center, the samples are assigned to the nearest cluster. K-means clustering can be used to discover potential structures or patterns in data.

In addition, algorithms and models such as neural networks (Figure 3), deep learning, and ensemble learning have also been widely applied in drug development. The selection of these algorithms and models depends on factors such as specific data characteristics, task requirements, and model performance.

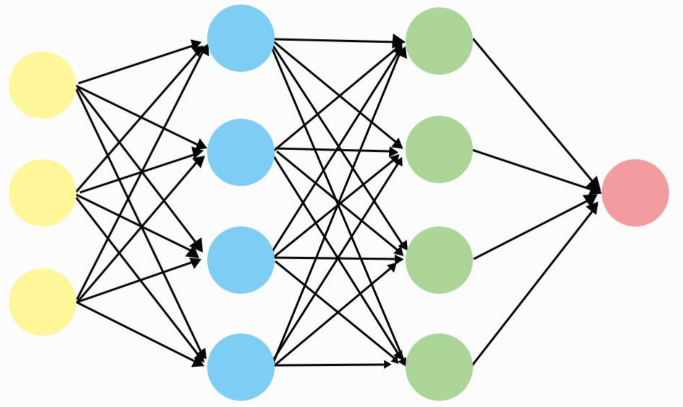

Figure 3 AI Neural Network Model

Note: Yellow: Input data; Blue: For hidden layers; Green: Process and target another hidden layer; Red: Make modifications to generate final output

3.3 Optimizing the model to meet the needs of drug screening

In order to ensure the accuracy and efficiency of machine learning models in drug screening, a series of targeted optimization work is needed. High quality data is the foundation of model performance, so steps such as removing noise, filling in missing values, standardizing or normalizing features, and handling imbalanced data are essential. Feature engineering is equally crucial, and selecting features closely related to drug activity can significantly improve the predictive ability of the model.

Select appropriate machine learning algorithms and models based on the specific needs of drug screening and the characteristics of the data. Different algorithms and models are suitable for different tasks and data types, so it is necessary to choose according to the actual situation. At the same time, it is also important to consider the interpretability and comprehensibility of the model in order to provide meaningful insights in the drug development process.

During the model construction process, by adjusting the hyperparameters of the model, such as learning rate, regularization strength, tree depth, etc., the performance of the model can be optimized. Using grid search, random search, or Bayesian optimization methods to find the optimal combination of parameters can help improve the prediction accuracy and stability of the model (Beaurivage et al., 2019). By dividing the dataset into training and validation sets (or multiple folds), training the model on the training set, and evaluating the model's performance on the validation set, a stable estimate of the model's performance can be obtained. This can better understand the generalization ability of the model and further optimize it based on it.

By combining the prediction results of multiple single models, integrated learning can improve the overall performance. In drug screening, ensemble learning methods such as random forests and gradient boosting trees can combine the advantages of multiple models to improve their accuracy and stability. Selecting features with clear biological significance or constructing interpretable models can help researchers understand the predictive results and underlying biological mechanisms of the models.

4 Candidate Drug Validation

4.1 Evaluation criteria and validation methods for candidate drugs

In the drug development process, the evaluation and validation of candidate drugs provide key measurement standards for the effectiveness and safety of drugs. The evaluation criteria mainly include pharmacodynamic evaluation, pharmacokinetic evaluation, and safety evaluation. Pharmacodynamic evaluation verifies whether a drug has the expected pharmacological effect through in vitro or in vivo experiments, ensuring that it matches the expected therapeutic effect. Pharmacokinetic evaluation focuses on the absorption, distribution, metabolism, and excretion processes of drugs in the body to predict their therapeutic effects and potential side effects. At the same time, by evaluating the potential toxicity, mutagenicity, carcinogenicity, and other aspects of the drug on the body, the safety of the drug is ensured.

To meet these evaluation criteria, the selection of validation methods is crucial. In vitro experiments are one of the commonly used verification methods, which provide a basis for preliminary evaluation of drug efficacy and safety by simulating the process of drug action in the body under laboratory conditions. Animal experiments further validate the performance of drugs in animal bodies, including evaluations of acute toxicity, long-term toxicity, and pharmacodynamics (Abbasnezhad et al., 2022). Clinical trials are a crucial step in verifying the efficacy and safety of drugs in practical applications, gradually validating the effectiveness and safety of drugs through multiple stages of trials.

The criteria and validation methods for evaluating candidate drugs are crucial steps in drug development, involving multiple techniques and analytical methods. Balbach and Korn (2004) discussed methods for evaluating candidate drugs in early drug development, emphasizing the importance of early evaluation of physicochemical parameters such as solubility, dissolution rate, hygroscopicity, lipophilicity, pKa, stability, polymorphism, and particle properties. They proposed a method that requires only up to 100 milligrams of drug substances for high-quality evaluation.

Benson et al. (2006) discussed the importance of validating cancer drug targets by demonstrating that specific therapeutic drugs are effective in clinical practice and acting through their designed targets, although it is advisable to declare early drug targets as "validated" before entering a comprehensive drug discovery program. Plenge et al. (2013) described the use of naturally occurring human mutations that affect the activity of specific protein targets, which can be used to estimate the potential efficacy and toxicity of drugs targeting such proteins, as well as establish causal relationships between targets and outcomes rather than reactive relationships. The above examples emphasize the multifaceted approach from physical and chemical evaluation to clinical evaluation and validation of biological analysis methods, which is crucial for ensuring the safety, efficacy, and ultimate clinical application of candidate drugs.

4.2 Experimental design and animal model selection

In drug development, experimental design and animal model selection not only directly affect the reliability and effectiveness of experimental results, but also determine the efficiency and cost of the drug development process. Experimental design requires comprehensive consideration of multiple factors, including experimental objectives, research questions, expected outcomes, and available resources and time. A good experimental design should be clear, repeatable, and able to accurately answer research questions. It is also necessary to consider factors such as variable control, sample size, data collection and analysis methods in the experiment to ensure the accuracy and reliability of the experimental results. For example, Silk et al. (2014) emphasized the importance of experimental design in model selection and parameter inference. Through experimental design and model selection framework based on random state space models, they demonstrated that the selected model may depend on the experiments performed. This indicates that the experimental design makes model selection a confidence criterion, but this is not necessarily related to the predictive power or correctness of the model.

The selection of animal models is also an indispensable part of experimental design. Different types of animals have different physiological and pathological characteristics. In the process of drug development, how to enhance the reliability and efficiency of research through precise experimental design and appropriate animal model selection. The correct selection and use of animal models are crucial for understanding disease mechanisms, evaluating drug safety and efficacy, and ultimately achieving successful drug development.

When selecting animal models, it is necessary to consider factors such as the characteristics of the disease, the genetic background of the animal, physiological and pathological reactions, and the operability of the experiment. For example, Alexander (2020) found that certain diseases may be more prominent in specific species of animals, or certain animals may be more sensitive to specific drug responses.

Singh and Seed (2021) discussed the importance of experimental animal models in drug discovery and development, particularly in understanding the origin, pathology, and development of safe and effective treatment and cure methods for human diseases. Although animal models are crucial in drug development, the low conversion rate of research results has led to many new drugs failing in clinical trials. This emphasizes the importance of selecting appropriate animal models and conducting precise experimental design in the early stages of drug development.

4.3 Analysis of experimental results and evaluation of the effectiveness of candidate drugs

The analysis of experimental results involves interpreting and interpreting experimental data, as well as objectively evaluating the efficacy of candidate drugs. This process not only requires researchers to have solid statistical and data analysis abilities, but also requires a deep understanding of the background of drug development and the original intention of experimental design. When analyzing experimental results, it is necessary to control the quality of the collected data to ensure its completeness and accuracy. Process and analyze data using appropriate statistical methods and data analysis tools. This may include descriptive statistics, analysis of variance, regression analysis, etc., to reveal the patterns and trends behind the data.

When evaluating the effectiveness of candidate drugs, it is necessary to comprehensively consider multiple indicators. The first is the pharmacodynamic indicator, which refers to the degree to which the drug affects the target biomolecule or cell in vitro or in vivo experiments. This is usually evaluated by comparing the differences between the drug treatment group and the control group. In addition, it is necessary to pay attention to pharmacokinetic indicators, understand the absorption, distribution, metabolism, and excretion processes of drugs in vivo, in order to predict their concentration changes and therapeutic effects in vivo.

In addition to the above indicators, safety assessment is also an indispensable part of evaluating the effectiveness of candidate drugs. This includes an evaluation of the potential toxicity, mutagenicity, carcinogenicity, and other aspects of the drug. Through animal experiments and clinical trials, the safety of drugs can be comprehensively evaluated, providing important basis for subsequent clinical applications and marketing. When evaluating the effectiveness of candidate drugs, attention should also be paid to the reliability and reproducibility of experimental results. This requires researchers to follow scientific principles and norms in experimental design and data analysis, ensuring the accuracy and credibility of experimental results. Multiple repeated experiments are required to verify the stability and consistency of drug efficacy.

5 Case Studies

5.1 Analysis of successful cases of AI based drug screening

In recent years, artificial intelligence (AI) technology has made significant progress in the field of drug screening, greatly accelerating the development process of new drugs. Here is a successful case of AI based drug screening:

Case name: AlphaFold helps to develop COVID-19 drugs

Case background: At the beginning of 2020, COVID-19 (SARS CoV-2) broke out in the world, and effective drugs and vaccines are urgently needed to deal with this global health crisis. The traditional drug development process is time-consuming and labor-intensive, so researchers have begun to explore the use of AI technology to accelerate the process of drug screening and design.

AI technology application: In this case, AlphaFold is a deep learning based protein structure prediction tool developed by DeepMind in the UK. It can predict the three-dimensional structure of proteins by analyzing amino acid sequences. This technology has played a key role in the research and development of COVID-19 drugs.

R&D process: The researchers first used AlphaFold to predict the three-dimensional structure of the main protease (Mpro) of COVID-19. Subsequently, they utilized this structural information to screen candidate drugs that may bind to Mpro from known drug databases. After further experimental verification, they successfully found several candidate drugs with the activity of inhibiting COVID-19.

Case achievement: Based on the AI drug screening method of AlphaFold, scientific researchers screened a variety of potential candidate drugs in a short time, providing strong support for drug research and development of COVID-19 (Jamilloux et al., 2020). These methods not only shorten the drug development cycle, but also reduce research and development costs, making important contributions to the global fight against the epidemic.

5.2 Key factors and application value for the success of the above cases

The key factors for the success of the above cases mainly include advanced AI technology, large-scale computing power, interdisciplinary cooperation, and rapid data acquisition and integration. AlphaFold, as an advanced deep learning tool, can accurately predict the three-dimensional structure of proteins, providing an important foundation for drug screening (Jamilloux et al., 2020). In addition, large-scale computing power ensures the efficient operation of AI models, while interdisciplinary collaboration promotes close collaboration among researchers in different fields. The rapid data acquisition and integration capabilities enable researchers to quickly respond to global health crises such as the pandemic, providing strong support for drug research and development.

The application value of this case is mainly reflected in the following aspects. Firstly, AI technology can greatly accelerate the drug development process, improve research and development efficiency, and shorten the time to market for new drugs. Secondly, by reducing the need for experimental verification, AI technology can help reduce the cost of drug development. In addition, by predicting and screening a large number of compounds, AI technology can improve the success rate of drug development and screen out more potential candidate drugs. Finally, in response to the global health crisis, AI technology has provided researchers with fast and efficient drug development methods, making important contributions to the global fight against crises such as the pandemic.

6 Discussion and Outlook

The current AI based drug screening process shows obvious advantages and disadvantages. The advantage lies in its speed and efficiency. AI technology can quickly process and analyze large amounts of data, significantly reducing drug screening time and improving research and development efficiency. The accuracy of AI algorithms is also higher, which can more accurately predict the interaction between drugs and targets, reducing the need for experimental verification (Neves et al., 2018). AI drug screening also helps to reduce the overall cost of drug development, achieving economic benefits by reducing the number of experiments and labor costs. However, its shortcomings cannot be ignored. The accuracy of AI models highly depends on the quality and quantity of input data. If there is bias or inadequacy in the data, it may lead to misleading screening results. At the same time, the interpretability of current AI models is insufficient, making it difficult for researchers to understand the working principle of the models and the basis for screening results. In addition, AI drug screening may also involve complex issues such as data privacy, intellectual property, and ethical review, which need to be handled with caution.

At present, technological challenges and data challenges are the two main challenges. It is necessary to improve the accuracy and reliability of AI models, especially when dealing with complex and diverse biological data. The data challenge lies in obtaining high-quality and diverse biological data, and solving the problems of data annotation and integration. Possible solutions to these challenges include continuous research and optimization of AI models, strengthening interdisciplinary cooperation to jointly promote the application of AI in drug development, and establishing a strict data governance system to ensure data quality and accuracy.

The AI based drug screening process is expected to present more development trends and potential impacts. Model integration and fusion may become an important direction in the future, improving the accuracy and efficiency of drug screening by integrating and fusing different types of AI models. The development of personalized healthcare will also benefit from the promotion of AI drug screening, which selects the most suitable drugs based on individual genetic, physiological, and pathological characteristics (Gorshkov et al., 2019). These development trends will also have potential impacts. The overall speed and efficiency of drug development are expected to be significantly improved, accelerating the speed of new drug launch. Medical costs may be reduced due to the widespread use of AI drug screening, benefiting more people. However, as the application of AI in drug development becomes increasingly widespread, it may also trigger more ethical and regulatory issues, requiring the development of corresponding policies and regulations to regulate it.

References

Abbasnezhad A., Salami F., and Mohebbati R., 2022, A review: Systematic research approach on toxicity model of liver and kidney in laboratory animals, Animal Models and Experimental Medicine, 5: 436-444.
https://doi.org/10.1002/ame2.12230
PMid:35918879 PMCid:PMC9610155

Alexander M., Schoeder C., Brown J., Smart C., Moth C., Wikswo J., Capra J., Meiler J., Chen W., and Madhur M., 2020, Predicting susceptibility to SARS-CoV-2 infection based on structural differences in ACE2 across species, The FASEB Journal, 34: 15946-15960.
https://doi.org/10.1096/fj.202001808R
PMid:33015868 PMCid:PMC7675292

Balbach S., and Korn C., 2004, Pharmaceutical evaluation of early development candidates "the 100 mg-approach", International Journal of Pharmaceutics, 275: 1-2, 1-12 .
https://doi.org/10.1016/j.ijpharm.2004.01.034
PMid:15081133

Beaurivage C., Naumovska E., Chang Y., Elstak E., Nicolas A., Wouters H., Moolenbroek G., Lanz H., Trietsch S., Joore J., Vulto P., Janssen R., Erdmann K., Stallen J., and Kurek D., 2019, Development of a gut-on-a-chip model for high throughput disease modeling and drug discovery, International Journal of Molecular Sciences, 20(22): 5661.
https://doi.org/10.3390/ijms20225661
PMid:31726729 PMCid:PMC6888156

Benson J., Chen Y., Cornell-Kennon S., Dorsch M., Kim S., Leszczyniecka M., Sellers W., and Lengauer C., 2006, Validating cancer drug targets, Nature, 441: 451-456.
https://doi.org/10.1038/nature04873
PMid:16724057

Burton P., Banner N., Elliot M., Knoppers B., and Banks J., 2017, Policies and strategies to facilitate secondary use of research data in the health sciences, International Journal of Epidemiology, 46: 1729-1733.
https://doi.org/10.1093/ije/dyx195
PMid:29025140 PMCid:PMC5837447

Cai J., Luo J., Wang S., and Yang S., 2018, Feature selection in machine learning: A new perspective, Neurocomputing, 300: 70-79.
https://doi.org/10.1016/j.neucom.2017.11.077

Costa C., Frampas C., Longman K., Palitsin V., Ismail M., Sears P., Nilforooshan R., and Bailey M., 2019, Paper spray screening and liquid chromatography/mass spectrometry confirmation for medication adherence testing: A two‐step process, Rapid Communications in Mass Spectrometry, 35(S2): E8553.
https://doi.org/10.1002/rcm.8553
PMid:31414505 PMCid:PMC8047880

Deng J., Yang Z., Samaras D., and Wang F., 2021, Artificial intelligence in drug discovery: applications and techniques, Briefings in bioinformatics, 23(1): bbab430.
https://doi.org/10.1093/bib/bbab430
PMid:34734228

Dogan T., 2018, UniProt: a worldwide hub of protein knowledge, Nucleic Acids Research, 47: 506-515.
https://doi.org/10.1093/nar/gky1049
PMid:30395287 PMCid:PMC6323992

Gorshkov K., Chen C., Marshall R., Mihatov N., Choi Y., Nguyen D., Southall N., Chen K., Park J., and Zheng W., 2019, Advancing precision medicine with personalized drug screening, Drug discovery today, 24(1): 272-278.
https://doi.org/10.1016/j.drudis.2018.08.010
PMid:30125678 PMCid:PMC6372320

Hessler G., and Baringhaus K., 2018, Artificial Intelligence in Drug Design, Molecules: A Journal of Synthetic Chemistry and Natural Product Chemistry, 23(10): 2520.
https://doi.org/10.3390/molecules23102520
PMid:30279331 PMCid:PMC6222615

Jamilloux Y., Henry T., Belot A., Viel S., Fauter M., Jammal T., Walzer T., François B., and Sève P., 2020, Should we stimulate or suppress immune responses in COVID-19? Cytokine and anti-cytokine interventions, Autoimmunity Reviews, 19: 102567-102567.
https://doi.org/10.1016/j.autrev.2020.102567
PMid:32376392 PMCid:PMC7196557

Karimi S., Wang C., Metke-Jimenez A., Gaire R., and Paris C., 2015, Text and data mining techniques in adverse drug reaction detection, ACM Computing Surveys (CSUR), 47: 1-39.
https://doi.org/10.1145/2719920

Kavakiotis I., Tsave O., Salifoglou A., Maglaveras N., Vlahavas I., and Chouvarda I., 2017, Machine learning and data mining methods in diabetes research, Computational and Structural Biotechnology Journal, 15: 104-116.
https://doi.org/10.1016/j.csbj.2016.12.005
PMid:28138367 PMCid:PMC5257026

Lei J., Zhou J., Zhao Y., Chen Z., Zhao P., Xie C., Ni Z., Yu T., and Xie J., 2021, Prediction of burn-up nucleus density based on machine learning, International Journal of Energy Research, 45: 14052-14061.
https://doi.org/10.1002/er.6660

Lin C., and Zhou X.X., 2022, Optimizing Drug Screening with Machine Learning, International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, pp.1-4.
https://doi.org/10.1109/ICCWAMTIP56608.2022.10016572

Mak K., and Pichika M., 2019, Artificial intelligence in drug development: present status and future prospects, Drug discovery today, 24(3): 773-780.
https://doi.org/10.1016/j.drudis.2018.11.014
PMid:30472429

Mitchell J., 2014, Machine learning methods in chemoinformatics, Wiley Interdisciplinary Reviews, Computational Molecular Science, (4): 468-481.
https://doi.org/10.1002/wcms.1183
PMid:25285160 PMCid:PMC4180928

Mohanty S., Rashid M., Mridul M., Mohanty C., and Swayamsiddha S., 2020, Application of artificial Intelligence in COVID-19 drug repurposing, Diabetes & Metabolic Syndrome, 14: 1027-1031.
https://doi.org/10.1016/j.dsx.2020.06.068
PMid:32634717 PMCid:PMC7332938

Neves B.J., Braga R.C., Melo-Filho C.C., Moreira-Filho J.T., Muratov E.N., and Andrade C.H., 2018, QSAR-Based virtual screening: advances and applications in drug discovery, Front. Pharmacol, 9: 1275.
https://doi.org/10.3389/fphar.2018.01275
PMid:30524275 PMCid:PMC6262347

Pan M., Xiang P., Yu Z., Zhao Y., and Yan H., 2019, Development of a high-throughput screening analysis for 288 drugs and poisons in human blood using Orbitrap technology with gas chromatography-high resolution accurate mass spectrometry, Journal of chromatography. A., 1587: 209-226 .
https://doi.org/10.1016/j.chroma.2018.12.022
PMid:30595433

Plenge R., Scolnick E., and Altshuler D., 2013, Validating therapeutic targets through human genetics,Nature Reviews Drug Discovery, 12: 581-594.
https://doi.org/10.1038/nrd4051
PMid:23868113

Ray S., 2019, A Quick Review of Machine Learning Algorithms, 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, pp.35-39.
https://doi.org/10.1109/COMITCon.2019.8862451

Silk D., Kirk P., Barnes C., Toni T., and Stumpf M., 2014, Model Selection in Systems Biology Depends on Experimental Design, PLoS Computational Biology, 10(6): E1003650.
https://doi.org/10.1371/journal.pcbi.1003650
PMid:24922483 PMCid:PMC4055659

Singh V., and Seed T., 2021, How necessary are animal models for modern drug discovery? Expert Opinion on Drug Discovery, 16: 1391-1397.
https://doi.org/10.1080/17460441.2021.1972255
PMid:34455867

Walke S., Nanvare K., Jamadade A., and Tembare R., 2023, A review paper on: clinical trials, International Journal For Multidisciplinary Research, 5(5): 7412.
https://doi.org/10.36948/ijfmr.2023.v05i05.7412

Walters W., and Barzilay R., 2021, Critical assessment of AI in drug discovery, Expert Opinion on Drug Discovery, 16: 937-947.
https://doi.org/10.1080/17460441.2021.1915982
PMid:33870801

Wilson A., Thabane L., and Holbrook A., 2003, Application of data mining techniques in pharmacovigilance, British Journal of Clinical Pharmacology, 57(2): 127-134 .
https://doi.org/10.1046/j.1365-2125.2003.01968.x
PMid:14748811 PMCid:PMC1884444

Wu G., Zhao T., Kang D., Zhang J., Song Y., Namasivayam V., Kongsted J., Pannecouque C., Clercq E., Poongavanam V., Liu X., and Zhan P., 2019, Overview of recent strategic advances in medicinal chemistry, Journal of Medicinal Chemistry, 62(21): 9375-9414.
https://doi.org/10.1021/acs.jmedchem.9b00359
PMid:31050421

Yang J., Li Y., Liu Q., Li L., Feng A., Wang T., Zheng S., Xu A., and Lyu J., 2020, Brief introduction of medical database and data mining technology in big data era, Journal of Evidence-Based Medicine, 13: 57-69.
https://doi.org/10.1111/jebm.12373
PMid:32086994 PMCid:PMC7065247

Yang X., Wang Y., Byrne R., Schneider G., and Yang S., 2019, Concepts of artificial intelligence for computer-assisted drug discovery, Chemical reviews, 119(18): 10520-10594.
https://doi.org/10.1021/acs.chemrev.8b00728
PMid:31294972