Review

DOI: 10.5582/irdr.2023.01111

The use of artificial intelligence in the treatment of rare diseases: A scoping review

Da He[1,§], Ru Wang[1,§], Zhilin Xu[2], Jiangna Wang[3], Peipei Song[4], Haiyin Wang[1,*], Jinying Su[5,*]

[1] Shanghai Health Development Research Center (Shanghai Medical Information Center), Shanghai, China;
[2] EYE & ENT Hospital of Fudan University, Shanghai, China;
[3] Jiangxi University of Chinese Medicine, Shanghai, China;
[4] Center for Clinical Sciences, National Center for Global Health and Medicine, Tokyo, Japan;
[5] Shanghai University of Traditional Chinese Medicine, Shanghai, China.

SUMMARY With the increasing application of artificial intelligence (AI) in medicine and healthcare, AI technologies have the potential to improve the diagnosis, treatment, and prognosis of rare diseases. Presently, existing research predominantly focuses on the areas of diagnosis and prognosis, with relatively fewer studies dedicated to the domain of treatment. The purpose of this review is to systematically analyze the existing literature on the application of AI in the treatment of rare diseases. We searched three databases for related studies, and established criteria for the selection of retrieved articles. From the 407 unique articles identified across the three databases, 13 articles from 8 countries were selected, which investigated 10 different rare diseases. The most frequently studied rare disease group was rare neurologic diseases (n = 5/13, 38.46%). Among the four identified therapeutic domains, 7 articles (53.85%) focused on drug research, with 5 specifically focused on drug discovery (drug repurposing, the discovery of drug targets and small-molecule inhibitors), 1 on pre-clinical studies (drug interactions), and 1 on clinical studies (information strength assessment of clinical parameters). Across the selected 13 articles, we identified total 32 different algorithms, with random forest (RF) being the most commonly used (n = 4/32, 12.50%). The predominant purpose of AI in the treatment of rare diseases in these articles was to enhance the performance of analytical tasks (53.33%). The most common data source was database data (35.29%), with 5 of these studies being in the field of drug research, utilizing classic databases such as RCSB, PDB and NCBI. Additionally, 47.37% of the articles highlighted the existing challenge of data scarcity or small sample sizes.

Keywords artificial intelligence, rare diseases, treatment

1. Introduction

Rare diseases are defined as illnesses with a low incidence rate. Different countries and regions have varying specifications for the incidence rates of rare diseases. In the European Union, diseases with an incidence rate lower than 1 in 2,000 are considered rare diseases (*1*); in the United States, rare diseases are defined as those affecting fewer than 200,000 individuals annually (or an incidence rate less than 1 in 1,500) (*2*); in Japan, rare diseases are specified as those affecting fewer than 50,000 individuals (or an incidence rate of 1 in 2,500) (*3,4*). According to the World Health Organization (WHO), rare diseases are those where the number of individuals affected constitutes 0.65‰ to 1‰ of the total population (*5*). With the continuous advancement in disease diagnostic technologies, the increasing subdivision in the field of diseases, and the yearly improvement in data statistics, new diseases have been continually identified or included since the definition of rare diseases was established, leading to an increase in the types of rare diseases. There are currently more than 7,000 known rare diseases globally (*6*), with an estimated accumulated prevalence of 3.5–5.9% and affecting more than 400 million people worldwide (*7,8*).

Due to the difficulty in diagnosing rare diseases, many are treated as common illnesses or remain undetected, suggesting that the actual number of patients is likely higher than statistical estimates. Furthermore, with advancements in diagnostic technologies and increased health literacy among populations, the number of individuals identified with rare diseases continues to expand (*9*).

Currently, due to unclear etiologies, the small number

of patients, among other reasons, rare disease patients face significant challenges compared to those with common diseases such as hypertension and diabetes. These challenges include difficulties in diagnosis, a lack of specific treatment techniques post-diagnosis, or the inability to afford available medications (*10,11*). Globally, only about 5% of rare diseases have effective treatment methods available. Even for rare diseases with existing treatment options, the cost of medications is often prohibitively expensive, imposing a substantial economic burden on both patients and society (*12*).

In recent years, due to the large volume and structural nature of data in the medical field, which align well with the needs of artificial intelligence, the development and application of artificial intelligence (AI) has permeated various domains within healthcare (*13-18*). In the field of rare diseases, the majority of AI applications pertain to disease screening, diagnosis, and prognosis, areas in which there is a substantial body of related literature (*19-21*), and where comprehensive review studies have already categorized and summarized the findings. While screening, diagnosis, and prognosis are undoubtedly important, for the vast number of patients with confirmed rare diseases, the need for treatment is more urgent (*22*). Currently, research and applied studies of AI in the treatment of rare diseases are relatively scarce, and there are no comprehensive review studies on the use of AI in the treatment of rare diseases.

In this scoping review, we explore scientific literature to investigate the application of AI in the treatment of rare diseases, and identify the key features used to train these AI models based on the pursued objectives. This includes specific groups of rare diseases that have been highlighted, therapeutic areas, key algorithms employed, data types used in models, as well as potential opportunities and challenges. These functionalities can support the future development of AI in the treatment of rare diseases.

2. The scoping review methods

This scoping review was followed to be designed and performed by the Preferred Reporting Items for Systematic reviews and meta-Analyses extension for scoping reviews (PRISMA-ScR) guideline (*23*).

2.1. Literature search

We conducted a systematic search across three databases for eligible articles: *i*) PubMed, *ii*) Web of Science, and *iii*) the Institute of Electrical and Electronics Engineers (IEEE) Xplore. The specific search terms used are detailed in Table 1 (Online Data, *http://www.irdrjournal.com/action/getSupplementalData.php?ID=179*).

Within the PubMed database, our primary search terms were "Rare Diseases/therapy"(Mesh) and "Artificial Intelligence"(Mesh). To comprehensively retrieve articles in the domain of artificial intelligence, we also included all associated entry terms related to artificial intelligence in PubMed. In addition, in order to extensively search for applications of rare diseases in the therapeutic area, we reviewed all 26 MeSH terms associated with the term "therapy" in the MeSH Database of PubMed. Based on this, we identified six terms – virtual reality exposure therapy, computer assisted therapy, radiotherapy, wearables, surgical robot, and drug therapy – as interventions for rare diseases. These terms were then combined with "Rare Diseases"(Mesh) for filtration.

In the IEEE Xplore, our search terms were ("All Metadata": Rare Diseases) combined with the aforementioned six therapeutic interventions. This combined search was further merged with the term ("All Metadata": artificial intelligence).

In the Web of Science database, our search terms were TS = (Rare Diseases) OR TS = (orphan disease), which were then combined with the six treatment methods. This combined search was further merged with the term TS = (Artificial Intelligence). Additionally, we incorporated the search terms "machine learning" and "deep learning".

2.2. Literature selection

After obtaining the potential articles, we conducted a screening of abstracts and full texts. The following criteria were employed to identify relevant literature concerning the use of artificial intelligence methods for rare disease treatment: *i*) Written in English, *ii*) Published or publicly available (*e.g.* conference proceedings) between January 1, 2010, and August 31, 2023. And exclusion criteria included: *i*) Not published in peer-reviewed journals or conference proceedings (*e.g.* preprints), *ii*) Not original research (*e.g.* reviews, editorials), *iii*) Not human patient data or scientific texts or publications (*i.e.*, articles using animal or simulated data were excluded), and *iv*) not rare disease topic.

2.3. Data extraction and synthesis

One reviewer (WJN) developed and extracted metadata for each article. Two reviewers (WR and HD) verified and refined the metadata for completeness. Upon selecting the relevant studies based on the eligibility criteria, the metadata extracted from these articles included: *i*) Publication year; *ii*) Country where the study was conducted (according to the senior author's affiliation); *iii*) Rare disease (diseases were specified using the Orphanet disorder name); *iv*) Rare disease group (according to the "preferential parent" of the disease as defined in the hierarchy of the Orphanet classification (*24*), the classification is based on the 34 disorder groups updated on Orphant's website in July 2023 (*25*); *v*) Therapeutic area; *vi*) Treatment method

(interventions used for rare diseases); *vii*) Purpose of using AI (objective of the study); *viii*) AI architecture (including types of algorithms such as deep learning and machine learning); *ix*) Data type; and *x*) Challenges. Table 2 (Online Data, *http://www.irdrjournal.com/action/getSupplementalData.php?ID=179*) summarizes the metadata elements.

3. Specific status description

The literature search identified a total of 407 unique records. After screening and assessing the articles for eligibility, 13 articles were included in the final analysis (The list of the selected 13 articles with their metadata elements extracted is available in Supplementary Table S1, *http://www.irdrjournal.com/action/getSupplementalData.php?ID=180*). Figure 1 displays the Preferred Reporting Items for Systematic reviews and meta-Analyses (PRISMA) flow diagram for article selection. After de-duplication, the literature search identified a total of 407 unique records, 281 articles were included in full-text screening. After screening and assessing the articles for eligibility, 13 articles were included in the final analysis.

3.1. Yearly publication trend

These studies originated from eight different countries. The majority of the publications (n = 5, 38.46%) came from the United States, followed by Italy (n = 2, 15.38%); Canada, Austria, France, India, Japan, and China each contributed one publication (n = 1, 7.69%).

Over the 13-year span from 2010 to 2023, articles meeting the criteria were found from 2017 to 2022. The number of articles rose from 1 in 2017 to 5 in 2021, and four articles were identified in 2022. Since the beginning of 2023 and as of August 31, 2023, no articles have been published that fulfill the selection criteria concerning the use of AI in the treatment of rare diseases.

3.2. Rare diseases and rare diseases group

Ten unique rare diseases were identified from the reviewed articles, with 3 articles focusing on general rare diseases rather than specific conditions. Table 3 presents the rare diseases identified in the selected articles, along with their corresponding orpha numbers. We classified the rare diseases identified from the reviewed articles into rare disease groups using the hierarchical structure defined by Orphanet. If an article aims to investigate specific characteristics of rare diseases that can be applied to general rare diseases. A rare disease can qualify for multiple groups, and the group as defined by the preferential parent in the classification hierarchy.

Of the 10 diseases, 3 (30%) had a prevalence of 1-9/100,000 patients, 2 (20%) had a prevalence of 1-9/1,000,000, 2 (20%) had a prevalence of 1–5/10,000, and 1 (10%) < 1/1,000,000, and for the remaining two diseases, Orphanet did not provide their prevalence.

The groups of rare diseases identified are illustrated in Figure 2 (Online Data, *http://www.irdrjournal.com/action/getSupplementalData.php?ID=179*). The most frequent group was rare neurologic disease, accounting for 5 out of 13 articles (38.46%). All other groups were represented by just one article each.

3.3. Therapeutic area and method

As illustrated in Figure 3 (Online Data, *http://www.irdrjournal.com/action/getSupplementalData.php?ID=179*), from the 13 articles selected based on our criteria, we identified four therapeutic areas. Among

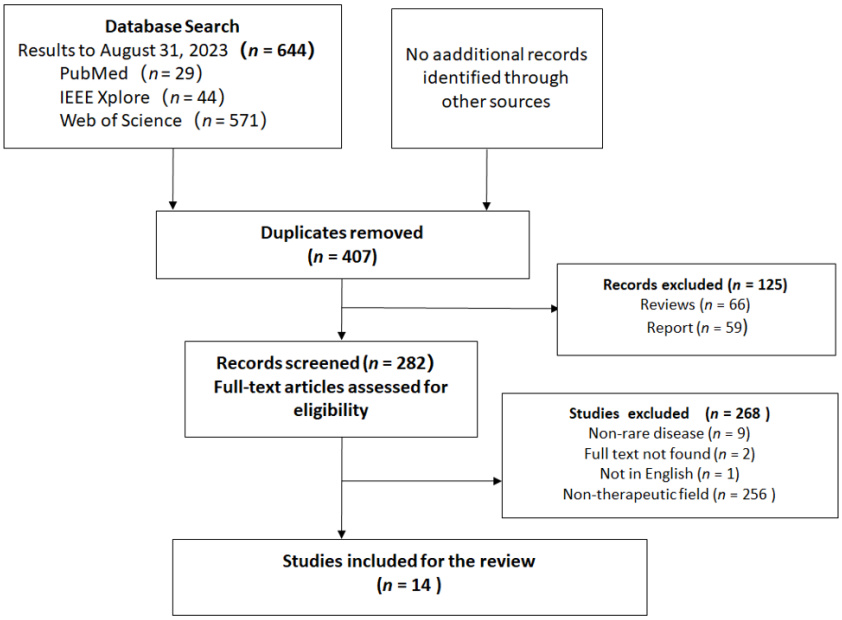

Figure 1. PRISMA flow diagram for literature selection.

Table 3. Rare diseases included in the selected literature

Rare disease	Orpha number (ORPHA)	Prevalence	Rare disease group (preferential parent)	Number of studies	Use cases (*Ref*)
General rare disease	NA	none	NA	3 (23.08%)	*26-28*
Metachromatic leukodystrophy	512	1–9/1,000,000	Rare neurologic disease	1 (7.69%)	*29*
Glycogen storage disease due to acid maltase deficiency	365	1–9/100,000	Rare genetic disease	1 (7.69%)	*30*
Canavan disease	141	Unknown	Rare neurologic disease	1 (7.69%)	*31*
Creutzfeldt-Jakob	204	< 1/1,000,000	Rare neurologic disease	1 (7.69%)	*32*
Pleural mesothelioma	50251	1–9/100,000	Rare neoplastic disease	1 (7.69%)	*33*
Alkaptonuria	56	1–9/1,000,000	Rare inborn errors of metabolism	1 (7.69%)	*34*
Acute Myeloid Leukaemi	519	1–5/10,000	Rare hematologic disease	1 (7.69%)	*35*
Congenital cataract microcornea with corneal opacity	289499	Unknown	Rare ophthalmic disorder	1 (7.69%)	*36*
Amyotrophic Lateral Sclerosis	803	1–9/100,000	Rare neurologic disease	1 (7.69%)	*37*
Canavan disease	141	Unknown	Rare neurologic disease	1 (7.69%)	*38*

these, 7 articles (53.85%) focused on drug research, 3 articles (23.08%) on precision medicine, 2 articles (15.38%) on health management, and 1 article (7.69%) on personalized services.

It's noteworthy that, of the 7 articles in the drug research area, 5 were focused on the area of drug discovery (*e.g.*, drug repurposing, the discovery of small-molecule drugs), 1 on pre-clinical studies (drug interactions), and 1 on clinical studies (information strength assessment of clinical parameters).

3.3.1. Drug research

Specifically, among the 5 articles focusing on drug discovery, 3 focused on the field of drug repurposing (*27-29*), that is, linking disease mechanisms with drug effects. Drug repurposing plays a significant role in the field of drug discovery. Foksinska *et al.* mentioned that identifying treatments for rare diseases is challenging due to limited understanding of disease mechanisms, small cohort sizes, inter-individual symptom variability, and little commercial incentive to develop new treatments. One promising therapeutic avenue was drug repurposing, where FDA-approved drugs were repurposed as new treatments (*26*). Challa *et al.* asserted that in its current state, ML (machine learning) was best leveraged in the field of drug repurposing to inform human "go/no-go" decision-making (*39*). Table 4 displays the overview of articles of drug research.

Cong *et al.* proposed a novel two-stage prediction approach for drug repurposing based on machine learning. This methodology clustered diseases based on gene expression patterns and evaluates drug efficacy through the reversibility of abnormal gene expression. It identified 22 drugs such as KM 00927, I-BET, alvocidib, and vorinostat, as candidates for repurposing, which had high efficacy against specific diseases like inclusion body myositis, polymyositis, and dermatomyositis (*27*). Sosa *et al.* proposed a literature-based knowledge graph embedding method for identifying drug repurposing opportunities in rare diseases. The method leveraged a large knowledge graph, the Global Network of Biomedical Relationships (GNBR), which integrated information from pharmacology, genetics, and pathology to generate drug repurposing hypotheses. The method achieved high performance on a gold-standard test set of known drug indications (AUROC = 0.89) and is capable of generating hypotheses for novel applications of existing drugs (*28*). Esmail *et al.* constructed an artificially induced whole-brain organoid platform (NEUBorg), which serves as an advanced iteration of the previously validated machine learning platform, DeepNEU (v6.2). Using NEUBorg, they generated artificially induced whole-brain organoids (aiWBO) simulations of metachromatic leukodystrophy (MLD) and provided a new method to evaluate factors related to MLD pathogenesis, disease progression and new potential treatment options. Utilizing this method, the authors identified 861 single and dual drug combinations as potential therapeutic targets for MLD. The study comprehensively summarizes the drug repurposing outcomes and the pharmaceuticals evaluated in the simulations (*29*).

One study focused on the discovery of drug targets. Esmail *et al.* constructed an AI platform, DeepNEU v3.6, to support target identification for Infantile Onset Pompe Disease (IOPD). This platform is capable of generating computer-simulated stem cells (aiPSC) and differentiated skeletal muscle cells (aiSkMC) models, both with and without the expression of Acid Alpha-Glucosidase (GAA). These simulations were validated using peer-reviewed results from existing literature to assess calcium homeostasis and mitochondrial function in IOPD patients. The authors employed aiSkMC IOPD simulations to identify known and novel biomarkers as well as potential therapeutic targets. Ultimately, the aiSkMC model for IOPD accurately predicted gene and phenotypic features reported in recent literature (*30*).

Another study focuses on the discovery of lead compounds for small-molecule inhibitors. Stecula *et al.* introduced a deep convolutional neural network, AtomNet, to identify lead compounds that inhibit Aspartate N-Acetyltransferase (ANAT), targeting the treatment of Canavan disease. The authors demonstrate AtomNet's capability to identify novel active scaffolds under challenging constraints, such as the scarcity or complete unavailability of target data, and thereby supporting early-stage drug discovery efforts (*31*).

Table 4. Overview of articles of drug research

Therapeutic area	Sub-area	Method	Algorithm	Product	Function and Result	*Ref.*
Drug research	Drug discovery	Drug repurposing	k-means, UMAP	A two-stage prediction approach for drug repurposing	This approach identified 22 drugs such as KM 00927, I-BET, alvocidib, and vorinostat, as candidates for repurposing, which had high efficacy against specific diseases like inclusion body myositis, polymyositis, dermatomyositis.	*27*
			UKG, UMAP	A literature-based knowledge graph embedding method	The method achieving high performance on a gold-standard test set of known drug indications (AUROC = 0.89) and is capable of generating hypotheses for novel applications of existing drugs.	*28*
			RNN, CM, SVM, GA	NEUBOrg	Could generate aiWBO simulations of MLD to evaluate Factors related to MLD pathogenesis, disease progression and new potential treatment options. The authors identified 861 single and dual drug combinations as potential therapeutic targets for MLD.	*29*
		Discovery of drug targets	RNN, CM, SVM, GA	DeepNEU v3.6	This platform is capable of generating aiPSC and aiSkMC models, both with and without the expression of GAA. The aiSkMC model for IOPD accurately predicted gene and phenotypic features reported in recent literature.	*30*
		Discovery of lead compounds for small-molecule inhibitors	CNN	AtomNet	to identify lead compounds that inhibit ANAT, targeting the treatment of Canavan disease.	*31*
	Pre-clinical research		LR, KNN, SVM, DT, RF, XGBoost	A model for studying the relationship between drugs used in the treatment of CJD and their impact on clinical parameters.	This model could suggest appropriate drugs upon the input of clinical parameters. The results indicated that Random Forest outperformed XGBoost, with an average accuracy of 98.39%.	*32*
	Clinical research		RF	Highlighted the informative power of clinical parameters in predicting initial response to treatment.	To predict the initial treatment response of patients with advanced/unresectable pleural mesothelioma and evaluate different clinical parameters (including gender, tissue type, BMI, smoking habits, number of packs/year and disease stage, etc.) and the correlation with response. The average AUC value of the model was 77.0%, the accuracy was 75%, the sensitivity was 74.8%, and the specificity was 83.3%.	*33*

In addition, one study focuses on pre-clinical research concerning drug interactions. Rajagopal *et al.* employed machine learning algorithms to develop a model for studying the relationship between drugs used in the treatment of Creutzfeldt-Jakob Disease (CJD) and their impact on clinical parameters. This model could suggest appropriate drugs upon the input of clinical parameters. The study evaluated various machine learning algorithms, such as Logistic Regression (LR), K-Nearest Neighbor (KNN), Decision Tree Classifier(DT), Support Vector Machine (SVM), Exterme Gradient Boosting (XGBoost), and RF. The results indicated that RF outperformed XGBoost, with an average accuracy of 98.39% (*32*).

Furthermore, one study highlighted the informative power of clinical parameters in predicting initial response to treatment. Massafra *et al.* used the RF algorithm combined with the sequential forward feature selection procedure to predict the initial treatment response of patients with advanced/unresectable pleural mesothelioma and evaluated different clinical parameters (including gender, tissue type, BMI, smoking habits, number of packs/year and disease stage, *etc.*) and the correlation with response. The average AUC (Area Under Curve) value of the model was 77.0%, the accuracy was 75%, the sensitivity was 74.8%, and the specificity was 83.3% (*33*).

3.3.2. Precision medicine

Drug repurposing can be difficult and requires depth of knowledge across multiple fields, which is complicated by the rapid pace of biomedical knowledge discovery.

To address these challenges, Foksinska *et al.* developed MediKanren, an artificial intelligence tool that used mechanistic insights into genetic diseases to identify treatment options to enable precision medicine for rare diseases. Utilizing knowledge graphs, mediKanren could effectively link all relevant literature and databases. The tool enabled a scalable process that has been used to help more than 500 rare disease families *(26*). Spiga *et al*. used machine learning algorithms, especially the RF algorithm, to establish a comprehensive digital ecosystem, AreciseKUre, which collected and analyzed data on the rare genetic disease Alkaptonuria (AKU), a digital platform aimed at creating a Precision Medicine Ecosystem (PME). AreciseKUre could determine the most suitable treatment method for AKU patients based on their quality of life (QoL) scores before and after medication. It represented a proof of principle study that could be applied to other rare diseases, allowing data management, analysis and interpretation (*27*). Licandro *et al.* proposed a novel method, WGAN-NN, for the computational quantification of Acute Myeloid Leukemia (AML) cancer cells in blood samples. WGAN-NN is a semi-supervised learning approach that embeds a Fully Connected Neural Network (FNN) within a Wasserstein Generative Adversarial Network (WGAN). The results show that the proposed semi-supervised WGAN embeddings are superior to PCA-NN and FNN (*35*). Table 5 displays the overview of articles of precision medicine.

3.3.3. Health management and personalized service

Long *et al.* introduced an artificial intelligence agent (AI agent), CC-Cruiser, that used deep learning for the diagnosis, risk stratification, and therapeutic recommendations for congenital cataracts. The AI agent was integrated with a cloud-based, multi-hospital collaboration platform aiming to improve disease management for patients with rare diseases such as congenital cataracts. Results indicated that CC-Cruiser accurately diagnosed congenital cataracts and provided treatment decisions across computational tests, web-based studies, needle-in-a-haystack tests, and multi-hospital clinical trials. The authors also demonstrated that the performance of the AI agent was comparable to that of individual ophthalmologists (*36*). Kmetzsch *et al.* presented a novel disease progression score using cross-sectional multimodal data to assess frontotemporal dementia and amyotrophic lateral sclerosis, a rare neurodegenerative disease. This framework leveraged neuroimaging and microRNA data to train supervised multimodal variational autoencoders to learn meaningful latent spaces that represented disease progression (*37*).

Chapron *et al.* discussed the impact of myotonic dystrophy type 1 (DM1) on the quality of life and daily activities of affected individuals in Quebec, Canada. To mitigate the effects of DM1, the authors proposed a complete assistance system that provided training guidance and motivation, Acti-DM1, which identified movement-related activities and monitored each exercise performed during training sessions (*38*). Table 6 displays the overview of articles of health management and personalized service.

3.4. Purpose of using AI and algorithms

In the 13 articles, we identified a total of 20 different algorithms (multiple algorithms were often studied in a single article. One article did not specify a particular algorithm but utilized the AI platform medKanren, which we categorized as "other"). The total count of algorithms was 32. The most frequently used algorithm was RF, accounting for 4 instances (12.50%); followed by SVM with 3 instances (9.37%); Subsequent to these were DT, KNN, Uniform Manifold Approximation and Projection (UMAP), Generative evolutionary algorithms (GA), Confusion Matrix (CM), Recurrent Neural Network (RNN), and Convolutional Neural Networks (CNN), each with 2 instances (6.25%); The algorithms k-means, WGAN, LR, XGBoost, FNN, Uncertain Knowledge Graph (UKG), Multilayer Perceptron (MLP), Naive Bayes (NB), Principal Component Analysis (PCA), Variational Autoencoder (VAE); and "Other" accounted for 1 instance (3.13%) (Figure 4, Online Data, *http://www.irdrjournal.com/action/getSupplementalData.php?ID=179*).

In the 13 articles, we identified 15 purposes for the use of AI in rare disease treatment, categorized into 4 types (a single article might have more than one

Table 5. Overview of articles of precision medicine

Therapeutic area	Algorithm	Product	Function and Result	(*Ref*)
Precision medicine	Other	MediKanren	MediKanren could effectively link all relevant literature and databases. The tool enabled a scalable process that has been used to help more than 500 rare disease families	*26*
	RF	AreciseKUre	AreciseKUre could determine the most suitable treatment method for AKU patients based on their quality of life (QoL) scores before and after medication.	*34*
	WGA, PCA, FNN	WGAN-NN	To calculate the number of AML cancer cells in blood samples. The results show that the proposed semi-supervised WGAN embeddings are superior to PCA-NN and FNN.	*35*

Table 6. Overview of articles of health management and personalized service

Therapeutic area	Algorithm	Product	Function and Result	(*Ref*)
Health management	CNN	CC-Cruiser	CC-Cruiser accurately diagnosed congenital cataracts and provided treatment decisions across computational tests, web-based studies, needle-in-a-haystack tests, and multi-hospital clinical trials. The authors also demonstrated that the performance of the AI agent was comparable to that of individual ophthalmologists.	*36*
	VAE	A novel disease progression score using cross-sectional multimodal data to assess frontotemporal dementia and amyotrophic lateral sclerosis.	This framework leveraged neuroimaging and microRNA data to train supervised multimodal variational autoencoders to learn meaningful latent spaces that represented disease progression.	*37*
Personalized service	RF, DT, SVM, KNN, NB, MLP	Acti-DM1	The system could identify movement-related activities and monitored each exercise performed during training sessions.	*38*

objective). The most prevalent purpose was to improve the performance of analytical tasks (*e.g.* improving model performance), accounting for 9 articles (53.33%); Proof-of-concept (*e.g.* analyzing the impact of drugs on clinical parameters) accounting for 3 articles (20.00%); To address the challenge of data scarcity was the goal in 3 articles (18.75%); and to reduce manual effort (*e.g.* leveraging AI tools for virtual drug screening to diminish the time and cost of manual experimentation) was highlighted in 1 article (6.67%) (Figure 5, Online Data, *http://www.irdrjournal.com/action/getSupplementalData.php?ID=179*).

3.5. Data type

Over the 13 articles, we identified a total of 7 distinct data sources, cumulatively accounting for 18 data points (Note that each paper might reference more than one data type). The predominant source of data was derived from databases, represented in 6 articles (35.29%). Following this, image data, exemplified by CT scan datasets, was cited in 3 articles (17.65%). Omics data (*e.g.* genomics, proteomics), demographic data (*e.g.* age), and functional test data (*e.g.* pulse, blood pressure, and other physiological metrics) were each represented in 2 articles (11.76%). Both literature-derived data and experimentally obtained data were cited in a single article each, contributing to 5.88% of the data sources.

It is noteworthy that within the data sourced from databases, 5 articles pertained specifically to the domain of drug development, with databases such as RCSB PDB and NCBI – quintessential repositories in the drug development landscape – being employed. Meanwhile, 2 articles employed database-derived data in the context of precision therapy (*e.g.*, ApreciseKUre) (Figure 6, Online Data, *http://www.irdrjournal.com/action/getSupplementalData.php?ID=179*).

3.6. Challenges in using the treatment of AI for rare diseases

We classified the challenges in using deep learning for rare disease research reported by the authors into 8 categories (Note that each paper might reference more than one challenge, and if no challenges were highlighted, they were labeled as "None"). Among the identified 19 challenges in applying AI in the field of rare diseases treatment, 10 articles (47.37%) mentioned the primary challenge is lack of sufficient data, indicating that the amount of available data has become the foremost challenge in using AI for rare disease research.

Enhancing interpretability stands as an imperative when applying AI in healthcare and medicine (*40,41*). While a multitude of studies have achieved commendable performance in their target analysis tasks, certain limitations warrant attention. Specifically, 15.76% (n = 3) of the articles stated that there's a need to have models with better interpretability. Additionally, 10.53% (n = 2) of the papers highlight the constrained availability of model options, indicating a pertinent need for diversified algorithms. The necessity for wet-lab validation within the realm of pharmaceutical research, signifies that future endeavors must bolster the reliability of AI methodologies. Another article highlighted that computational capabilities of wrist-worn devices used in their study were limited, necessitating a reduction in feature dimensions for machine learning algorithms (*38*). One article articulated that the dearth of suitable therapeutic options and limited acquired knowledge could potentially inhibit the utility of AI in the domain of rare diseases (*26*) (Figure 7, Online Data, *http://www.irdrjournal.com/action/getSupplementalData.php?ID=179*).

4. The status quo and problems of AI application in the treatment of rare diseases

AI has been extensively utilized in various types of fundamental research to date. With the advent and rapid iteration of large language models such as ChatGPT, AI's involvement has emerged in the therapeutic domains of

many diseases (*42-46*). This study provides a review of research applying AI technologies in the treatment of rare diseases, aiming to understand the current global contributions of AI in the field of rare disease therapy.

4.1. The application of AI in the treatment of rare diseases emerged later and is relatively scarce

Although the concepts of rare diseases and artificial intelligence were introduced quite some time ago, research on the application of AI in the treatment of rare diseases did not emerge until 2017. Over time, the number of studies has not seen a substantial increase, unlike in the areas of rare disease prevention, screening, and prognosis (*47,48*).

In terms of the countries involved in the research, the United States leads in applying AI to the treatment of rare diseases, followed by Italy, Canada, Australia, and France – countries that are more advanced in biomedicine and artificial intelligence – as well as three Asian countries: Japan, China, and India. However, overall, the number of studies on the application of AI in the treatment of rare diseases is relatively low across all countries.

From the perspective of the disease categories included in the literature, a relatively higher proportion of studies focuses on neurological aspects, suggesting that research on neurological drugs and therapies is more prevalent among various rare disease treatments.

The complete lifecycle of a disease includes prevention, screening, diagnosis, treatment, and rehabilitation. In this process, compared to diagnosis, screening, and prognosis of rare diseases, studies on the application of AI in their treatment are noticeably less. In a 2020 scoping review on the application of machine learning in the field of rare diseases, of the 211 included studies, 40.8% used machine learning for diagnosis, 38.4% for prognosis, and only 4.7% for treatment (*48*). A systematic review included studies on the use of machine learning for diagnosis and prognosis, published in July 2023, and 22 studies were included (*19*). The timeframe for literature retrieval in this study was similar to that of our research, yet the number of studies incorporated was 1.6 times greater than those included in our study.

Of the 13 articles included in this study, more than half aimed to improve the performance of analytical models, while only 3 addressed the significant data scarcity issue in rare diseases. This highlights the considerable challenges in applying AI technology to treatment of rare diseases. Besides, prevention and screening of rare diseases aim for earlier and faster treatment of symptoms. For the large number of patients already diagnosed with rare diseases and for those who cannot avoid being born with rare conditions in the future, the need for effective treatment is more urgent than prevention and screening. Therefore, it is recommended that global efforts further prioritize research on rare disease medications and treatment methods, fully leveraging advanced algorithms and models such as AI to accelerate the development of corresponding therapies.

4.2. Drug research is the primary domain for AI application in the treatment of rare diseases

In the field of innovative drug research, the 'Double Ten Law' - indicating that it takes 10 years and an investment of one billion US dollars to develop a single innovative drug - is widely recognized globally due to the long development cycle, high investment, and substantial risk associated with drug research (*49,50*). For rare diseases, the development of innovative drugs is even more challenging due to their low incidence rates, complex pathogenic mechanisms, insufficient numbers of subjects for clinical trials, and lower expected product sales, all of which reduce the commercial incentive for drug development compared to common diseases (*51-54*).

The role of artificial intelligence in drug development primarily includes: First, target identification – AI can analyze large-scale data, such as genomic and proteomic data, to identify potential drug targets and predict which targets have a higher likelihood of success. Second, compound screening – following target identification, AI can be used to screen large databases to find compounds that may interact with the target and predict which compounds are more likely to be successful, allowing researchers to prioritize these molecules for further testing. Third, compound optimization – AI can also be used to optimize the chemical structure of compounds, predicting how changes in chemical structure can affect the properties of the compound, such as its ability to bind to the target and potential toxicity, thus enhancing its efficacy and safety. Fourth, clinical trial design – AI can analyze past trial data and predict which patients are most likely to benefit from a specific drug, thereby conducting more effective clinical trials. Fifth, AI can help scientists more quickly hypothesize and validate potential treatments for rare diseases (*55-57*).

According to the review results, one of the main strategies to address the challenges in drug development for rare diseases is repurposing of existing drugs, where AI algorithms, including deep learning and knowledge graph technologies, can play a significant role. By deepening the understanding of drug targets and compound synthesis through AI, existing drugs can be repurposed for new indications or used in combination to treat rare diseases. Additionally, using machine learning, deep learning, and other intelligent algorithms to predict potential gene targets for rare diseases, accurately identify possible rare disease drugs, and suggest appropriate medications for different patients are effective methods to address the challenges in drug development for rare diseases (*58*).

With the assistance of AI, the research time for therapeutic drugs for rare diseases can be significantly reduced, and the efficiency of development greatly enhanced, thereby lowering costs and increasing the likelihood of successful drug research.

4.3. AI contributes to enhancing the efficiency of precision treatment and health management for patients with rare diseases

Based on the understanding of some pathogenic genes responsible for rare diseases, AI can be utilized for its powerful capability in linking, expanding, learning, and computing large datasets. By integrating algorithms with specific databases, it is possible to conduct exploratory drug pairing or testing at the genetic and molecular level, thereby precisely identifying treatment methods for rare diseases. Additionally, AI can identify subgroups of patients most likely to respond to specific treatments, thereby increasing the effectiveness of therapy and reducing the risk of adverse events. Furthermore, AI can circumvent the inefficient resource allocation inherent in traditional human-led health management, widely disseminating the expertise of highly skilled physicians and enabling long-term health management of patients (*59*).

4.4. The lack of high-quality data and the difficulty in constructing and interpreting complex models are the primary challenges in applying AI to rare disease treatment

Although AI demonstrates immense potential in drug repurposing, precision treatment, health management, and personalized services for rare diseases, numerous challenges need to be addressed before its full potential can be realized, given its relatively recent emergence.

First, the lack of high-quality data presents a significant challenge. AI algorithms depend on large volumes of high-quality data for learning and making accurate predictions. However, the low incidence of rare diseases results in a smaller data pool, which can impact the development of AI algorithms. Additionally, the quality of this limited data is subject to various factors, potentially limiting the effectiveness of AI algorithms. In drug discovery, this implies the necessity of acquiring more high-quality data regarding drug targets, chemical structures, and biological pathways.

Second, the complexity in model construction and interpretation poses a challenge. The high complexity of biological systems necessitates the development of algorithms that can predict interactions between drugs and biological systems and generate novel candidate drugs that are both safe and effective. While there are existing methods to extract feature importance from black-box models or to visualize some operational mechanisms of models, comprehensively understanding and interpreting the inner workings and reasoning processes of complex models remain significant challenges.

4.5. Ethical considerations in the use of AI for rare disease treatment

The application of AI in drug discovery also faces ethical and regulatory challenges, including ensuring the safety and efficacy of drugs designed by AI and avoiding unintended consequences. Furthermore, it is crucial to ensure that the use of AI is responsible and adheres to ethical standards (*60-62*).

4.6. Limitations

There are some limitations to our study. Primarily, the scope of investigation was restricted to research implementing artificial intelligence methodologies within the domain of rare disease treatment. Despite the paucity of publications on this specific subject, articles were not excluded based on the caliber of their publishing medium (*e.g.* Journal Impact Factor, JIF), potentially introducing a degree of bias into the review outcomes. Furthermore, the exhaustive list of 7,000+ rare disease instances reported by Orphanet was not entirely leveraged as search terms, potentially precluding the identification of pertinent articles on a broader spectrum of rare diseases. A final constraining factor was the exclusive consideration of articles penned in English, possibly omitting germane articles written in other languages.

5. Conclusion

First, the treatment of rare diseases is equally as important as prevention and screening. However, current AI research in rare diseases is disproportionately focused on drug development, with less emphasis on treatment. There is a need for continued research in drug repurposing, target discovery, hit compound identification, precision treatment, health management, and personalized services. Second, the development and interpretability of algorithms remain one of the main barriers in the treatment of rare diseases. There is potential for the application of large language models in the treatment of rare diseases. Last, to address the issue of insufficient data for rare disease treatment, it is worth considering the expansion of data sources beyond existing databases to include real-world data, especially the establishment of global databases for individual diseases.

Funding: None.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Julkowska D, Austin CP, Cutillo CM, Gancberg D, Hager C, Halftermeyer J, Jonker AH, Lau LPL, Norstedt I, Rath A, Schuster R, Simelyte E, van Weely S. The importance of international collaboration for rare diseases research: A European perspective. Gene Ther. 2017; 24: 562-571.
2. Vickers PJ. Challenges and opportunities in the treatment of rare diseases. Drug Discov World. 2013; 14:9-16.
3. Tada M, Hirata M, Sasaki M, Sakate R, Kohara A, Takahashi I, Kameoka Y, Masui T, Matsuyama A. The rare disease bank of Japan: Establishment, current status and future challenges. Human Cell. 2018; 31:183-188.
4. Ninomiya K, Okura M. Nationwide comprehensive epidemiological study of rare diseases in Japan using a health insurance claims database. Orphanet J Rare Dis. 2022; 17:1-13.
5. Song P, Gao J, Inagaki Y, Kokudo N, Tang W. Intractable and rare diseases research in Asia. Biosci Trends. 2012; 6:48-51.
6. Melnikova I. Rare diseases and orphan drugs[J]. Nature Reviews Drug Discovery. 2012; 11(4):267.
7. Hatem S, Long JC, Best S, Fehlberg Z, Nic Giolla Easpaig B, Braithwaite J. Mobile apps for people with rare diseases: Review and quality assessment using mobile app rating scale. J Med Internet Res. 2022; 24:e36691.
8. Nguengang WS, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, Le Cam Y, Rath A. Estimating cumulative point prevalence of: Analysis of the Orphanet database, Eur J Hum Genet. 2020; 28:165-173.
9. Marwaha S, Knowles JW, Ashley EA. A guide for the diagnosis of rare and undiagnosed disease: Beyond the exome. Genome Med. 2022; 14:23.
10. AAMC. Rare diseases difficult to diagnose, cures hard to come by. *https://www.aamc.org/news/rare-diseases-difficult-diagnose-cures-hard-come* (accessed November 11, 2023).
11. Ragni MV, Moore CG, Bias V, Key NS, Kouides PA, Francis CW. Challenges of rare disease research: Limited patients and competing priorities. Haemophilia. 2012; 18:e192-e194.
12. Kaufmann P, Pariser AR, Austin C. From scientific discovery to treatments for rare diseases – the view from the national center for advancing translational sciences – office of rare diseases research. Orphanet J Rare Dis. 2018; 13:196.
13. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019; 17:195.
14. Lee DH, Yoon SN. Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. Int J Environ Res Public Health. 2021; 18:271.
15. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: Transforming the practice of medicine. Future Healthc J. 2021; 8:e188-e194.
16. Aung YYM, Wong DCS, Ting DSW. The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare. British Med Bull. 2021; 139:4-15.
17. Van Hartskamp M, Consoli S, Verhaegh W, Petkovic M, van de Stolpe A. Artificial intelligence in clinical health care applications. Interactive J Med Res. 2019; 8:e12100.
18. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res. 2019; 21:e13659.
19. Roman-Naranjo P, Parra-Perez AM, Lopez-Escamez JA. A systematic review on machine learning approaches in the diagnosis and prognosis of rare genetic diseases. J Biomed Inform. 2023; 143:104429.
20. Caballé-Cervigón N, Castillo-Sequera JL, Gómez-Pulido JA, Gómez-Pulido JM, Polo-Luque ML. Machine learning applied to diagnosis of human diseases: A systematic review. Appl Sci. 2020; 10:5135-5161.
21. Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda. J Ambient Intell Humaniz Comput. 2023; 14:8459-8486.
22. Vickers PJ. Challenges and opportunities in the treatment of rare diseases. Drug Discov World Spring. 2013; 14:9-16.
23. Tricco AC, Lillie E, Zarin W, *et al.* PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018; 169:467-473.
24. Orphanet. Procedural document: Orphanet nomenclature and classification. *https://www.orpha.net/orphacom/cahiers/docs/GB/eproc_disease_inventory_R1_Nom_Dis_EP_04.pdf* (accessed November 11, 2023).
25. Orphanet Knowledge base release of July 2023. *https://www.orphadata.com/classifications/* (accessed November 11, 2023).
26. Foksinska A, Crowder CM, Crouse AB, Henrikson J, Byrd WE, Rosenblatt G, Patton MJ, He K, Tran-Nguyen TK, Zheng M, Ramsey SA, Amin N, Osborne J, Might M. The precision medicine process for treating rare disease using the artificial intelligence tool mediKanren. Front Artif Intell. 2022; 5:910216.
27. Cong Y, Shintani M, Imanari F, Osada N, Endo T. A new approach to drug repurposing with two-stage prediction, machine learning, and unsupervised clustering of gene expression. OMICS. 2022; 6:339-347.
28. Sosa DN, Derry A, Guo M, Wei E, Brinton C, Altman RB. A literature-based knowledge graph embedding method for identifying drug repurposing opportunities in rare diseases. Pac Symp Biocomput. 2020; 25:463-474.
29. Esmail S, Danter WR. Artificially induced pluripotent stem cell-derived whole-brain organoid for modelling the pathophysiology of metachromatic leukodystrophy and drug repurposing. Biomedicines. 2021; 9:440.
30. Esmail S, Danter W R. DeepNEU: Artificially induced stem cell (aiPSC) and differentiated skeletal muscle cell (aiSkMC) simulations of infantile onset POMPE disease (IOPD) for potential biomarker identification and drug discovery. Front Cell Dev Biol. 2019; 7:325.
31. Stecula A, Hussain MS, Viola RE. Discovery of novel inhibitors of a critical brain enzyme using a homology model and a deep convolutional neural network. J Med Chem. 2020; 63:8867-8875.
32. Rajagopal, Ganesh B, Arock M. Application of machine learning techniques for study of drug interactions using clinical parameters for Creutzfeldt-Jakob disease. 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), IEEE. 2021; pp:415-420.
33. Massafra R, Catino A, Perrotti PMS, Pizzutilo P, Fanizzi A, Montrone M, Galetta D. Informative power evaluation of clinical parameters to predict initial therapeutic response in patients with advanced pleural mesothelioma: A machine learning approach. J Clin Med. 2022; 11:1659.
34. Spiga O, Cicaloni V, Visibelli A, Davoli A, Paparo MA, Orlandini M, Vecchi B, Santucci A. Towards a precision

medicine approach based on machine learning for tailoring medical treatment in alkaptonuria. Int J Mol Sci. 2021; 22:1187.
35. Licandro R, Schlegl T, Reiter M, Diem M, Dworzak M, Schumich A, Langs G, Kampel M. WGAN latent space embeddings for blast identification in childhood acute myeloid leukaemia. 2018 24th International Conference on Pattern Recognition (ICPR), IEEE. 2018; pp:3868-3873.
36. Long E, Lin H, Liu Z, *et al*. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nat Biomed Eng. 2017; 1:0024
37. Kmetzsch V, Becker E, Saracino D, Rinaldi D, Camuzat A, Le Ber I, Colliot O. Disease progression score estimation from multimodal imaging and microRNA data using supervised variational autoencoders. IEEE J Biomed Health Inform. 2022; 26:6024-6035.
38. Chapron K, Lapointe P, Lessard I, Lessard I, Darsmstadt-Bélanger H, Bouchard K, Gagnon C, Lavoie M, Duchesne E, Gaboury less S. Acti-DM1: Monitoring the activity level of people with myotonic dystrophy type 1 through activity and exercise recognition. IEEE Access. 2021; 9:49960-49973.
39. Challa AP, Zaleski NM, Jerome RN, Lavieri RR, Shirey-Rice JK, Barnado A, Lindsell CJ, Aronoff DM, Crofford LJ, Harris RC, Alp Ikizler T, Mayer IA, Holroyd KJ, Pulley JM. Human and machine intelligence together drive drug repurposing in rare diseases. Fron Genet. 2021; 12:707836.
40. Ahmad MA, Eckert C, Teredesai A. Interpretable machine learning in healthcare. Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics. 2018; pp:559-560.
41. Vellido A. The importance of interpretability and visualization in machine learning for applications in medicine and health care. Neural Computing Applications. 2020; 32:18069-18083.
42. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell. 2023; 6:1169595.
43. Wojtara M, Rana E, Rahman T, Khanna P, Singh H. Artificial intelligence in rare disease diagnosis and treatment. Clin Transl Sci. 2023; 16:2106-2111.
44. Flores AM, Demsas F, Leeper NJ, Ross EG. Leveraging machine learning and artificial intelligence to improve peripheral artery disease detection, treatment, and outcomes. Circ Res. 2021; 128:1833-1850.
45. Chen HY, Ge P, Liu JY, Qu JL, Bao F, Xu CM, Chen HL, Shang D, Zhang GX. Artificial intelligence: Emerging player in the diagnosis and treatment of digestive disease. World J Gastroenterol. 2022; 28:2152-2162.
46. Dilsizian SE, Siegel EL. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cardiol Rep. 2014; 16:441.
47. Lee J, Liu C, Kim J, Chen Z, Sun Y, Rogers JR, Chung WK, Weng C. Deep learning for rare disease: A scoping review. J of Biomed Inform. 2022; 135:104227.
48. Schaefer J, Lehne M, Schepers J, Prasser F, Thun S. The use of machine learning in rare diseases: a scoping review. Orphanet J Rare Dis, 2020; 15:145.
49. Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B. 2022; 12:3049-3062.
50. Kiriiri GK, Njogu PM, Mwangi AN. Exploring different approaches to improve the success of drug discovery and development projects: A review. Future J Pharmaceutical Sciences. 2020; 6:1-12.
51. Schieppati A, Henter JI, Daina E, Aperia A. Why rare diseases are an important medical and social issue. The Lancet. 2008; 371:2039-2041.
52. Stoller JK. The challenge of rare diseases. Chest. 2018; 153:1309-1314.
53. Groft SC, Posada de la Paz M. Preparing for the future of rare diseases. Adv Exp Med Biol. 2017; 1031:641-648.
54. Mohs RC, Greig NH. Drug discovery and development: Role of basic biological research. Alzheimers Dement (NY). 2017; 3:651-657.
55. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol. 2022; 23:40-55.
56. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discov Today. 2021; 26:80-93.
57. Mak KK, Pichika MR. Artificial intelligence in drug development: Present status and future prospects. Drug Discov Today. 2019; 24:773-780.
58. Chen W, Liu X, Zhang S, Chen S. Artificial intelligence for drug discovery: Resources, methods, and applications. Mol Ther Nucleic Acids. 2023; 31:691-702.
59. Romm EL, Tsigelny IF. Artificial intelligence in drug treatment. Annu Rev Pharmacol Toxicol. 2020; 60:353-369.
60. Keskinbora KH. Medical ethics considerations on artificial intelligence. J Clin Neurosci. 2019; 64:277-282.
61. Naik N, Hameed B M, Shetty D K, Swain D, Shah M, Paul R, Aggarwal K, Ibrahim S, Patil V, Smriti K, Shetty S, Rai BP, Chlosta P, Somani BK. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? Front Surg. 2022; 9:862322.
62. Safdar NM, Banja JD, Meltzer CC. Ethical considerations in artificial intelligence. European J Rdiol. 2020; 122:108768.

Received September 26, 2023; Revised November 28, 2023; Accepted December 22, 2023.

[§]These authors contributed equally to this work.
**Address correspondence to:*
Haiyin Wang, Shanghai Health Development Research Center (Shanghai Medical Information Center), No. 181, Xinbei Road, Minhang District, Shanghai 201199, China.
E-mail: wanghaiyin@shdrc.org

Jinying Su, Shanghai University of Traditional Chinese Medicine, No.1200 Cailun Road, Zhangjiang Hi-Tech Park, Pudong New Area, Shanghai 201203, China.
E-mail: szysujinying@163.com

Released online in J-STAGE as advance publication December 30, 2023.