Artificial intelligence in rare disease diagnosis and treatment

Magda Wojtara[1] | **Emaan Rana**[2] | **Taibia Rahman**[3] | **Palak Khanna**[4] | **Heshwin Singh**[5]

[1]Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA

[2]Department of Science, University of Western Ontario, London, Ontario, Canada

[3]Department of Medicine, David Tvildiani Medical University, Tbilisi, Georgia

[4]Department of Medicine, Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia

[5]Department of Biology, Stony Brook University, Stony Brook, New York, USA

Correspondence
Magda Wojtara, Department of Human Genetics, University of Michigan, 440 Church Street, Ann Arbor, MI 48109, USA.
Email: wojtaram@umich.edu

Abstract

Artificial intelligence (AI) utilization in health care has grown over the past few years. It also has demonstrated potential in improving the efficiency of diagnosis and treatment. Some types of AI, such as machine learning, allow for the efficient analysis of vast datasets, identifying patterns, and generating key insights. Predictions can then be made for medical diagnosis and personalized treatment recommendations. The use of AI can bypass some conventional limitations associated with rare diseases. Namely, it can optimize traditional randomized control trials, and may eventually reduce costs for drug research and development. Recent advancements have enabled researchers to train models based on large datasets and then fine-tune these models on smaller datasets typically associated with rare diseases. In this mini-review, we discuss recent advancements in AI and how AI can be applied to streamline rare disease diagnosis and optimize treatment.

INTRODUCTION

Health care is one of several industries being revolutionized by rapidly developing technologies known as artificial intelligence (AI) and machine learning (ML). AI is the capacity of machines to carry out operations that ordinarily need human intelligence, such as learning, problem-solving, and decision making. The ability for machines to learn from experience and enhance their performance without being explicitly programmed is provided by ML, a subset of AI.[1] These technological advancements are able to analyze massive volumes of information, identify trends, and arrive at decisions with greater speed and precision than ever before by emulating human intelligence. From enhancing patient outcomes to optimizing hospital operations, these technologies offer a wide range of uses in health care. AI and ML are currently being utilized in the healthcare industry to enhance medical image processing, disease prediction and prevention, and hospital operations.[2] By leveraging these technologies, healthcare providers can diagnose and treat patients more accurately and efficiently. Additionally, AI and ML can help physicians optimize and accelerate the timeline for the diagnosis, treatment, and management of diseases.[3] As these technologies continue to evolve, more innovative solutions will emerge that will hold the capability of

revolutionizing and improving healthcare for all. Figure 1 provides an overview of some emerging types of AI.

Individuals with rare diseases face numerous challenges, including late diagnosis and misdiagnosis, improper or no response to therapies, and a lack of accurate monitoring tools. Misdiagnosing rare diseases is a significant obstacle that can lead to worsened symptoms along with the development of other health problems, ultimately resulting in a decline in patient well-being. Furthermore, patients with rare diseases are frequently hospitalized and suffer long-term complications as therapies do not elicit the proper response or only have a partial effect that diminishes over time.[4] These challenges in the diagnosis and treatment of rare diseases affect more than 350 million people worldwide and create a substantial economic burden on the healthcare system as well as result in poor patient outcomes.[5] Therefore, it is imperative to develop a tool that can assist in early diagnosis, improve treatment effectiveness, and monitor conditions to improve care and reduce costs. One such tool is AI, which has, in the literature, shown benefits for common and rare disease diagnoses and treatments. AI has already been utilized in oncology, as it can predict survival time, recurrence risks metastasis, and therapy response among other key factors that influence prognosis as well as uses its computing power to create diagnosis systems and screen health records.[1] Recently, AI has undergone significant improvements because previous iterations allowing for its possible application for less common diseases which often have smaller datasets.

AI IN RARE DISEASE DIAGNOSIS

AI works as an accomplice in integrating and examining diversified data. Diagnostic decision support systems effectively assist the medical practitioner by providing a list of relevant differential diagnoses. These systems have previously been effectively utilized for a variety of well-known use cases. Recently, it has been harnessed for the early detection and diagnosis of coronavirus disease 2019 (COVID-19) through monitoring of demographic, clinical, and epidemiological characteristics of patients.[6] These systems are also useful for implementation for rare diseases (RDs). The RDs, also sometimes referred to as orphan diseases, can stand to benefit from quicker and more efficient diagnoses. Algorithms have been designed and are already used to compile networks and register information through patients on rare diseases to identify new cases.[6] For instance, a combination of brain function and
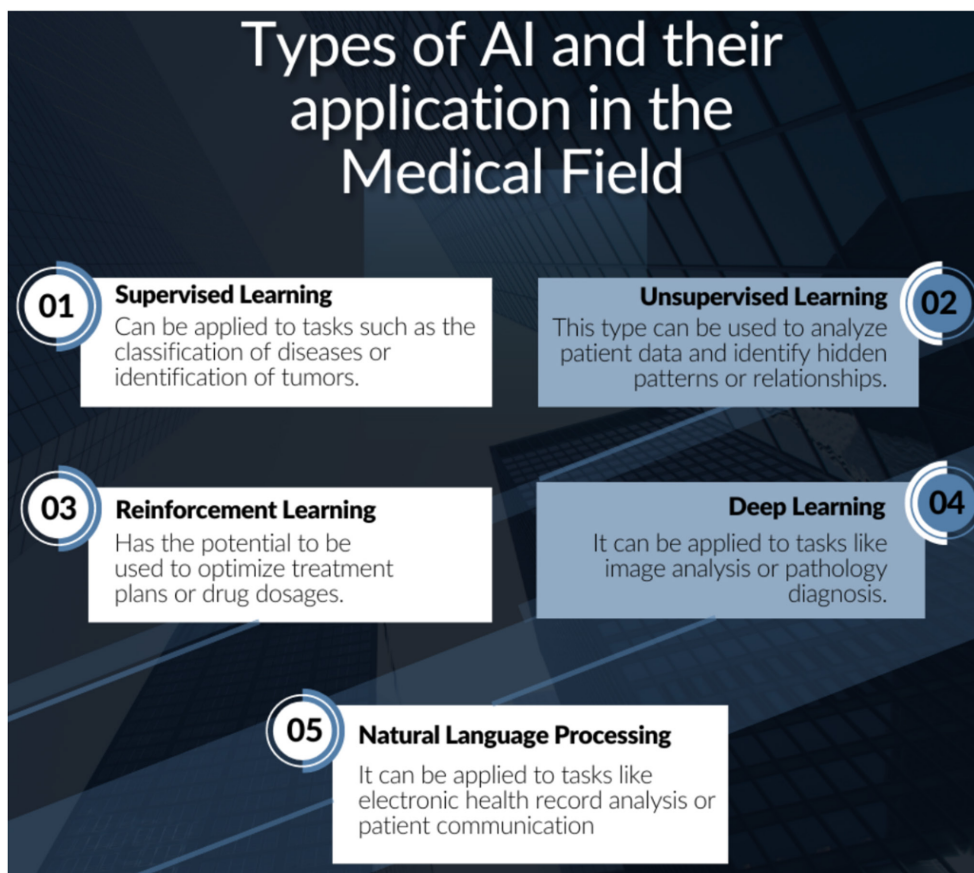

FIGURE 1 Types of artificial intelligence (AI) and their application in the medical field.

structural imaging data can be harnessed to determine whether a person with Huntington's disease (HD) will receive a clinical diagnosis within 5 years (pre-HD) or quantifiable assessments of oculomotor function preceding HD.[6] These use cases show promising potential for future utilization of AI in rare disease diagnoses.

Types of algorithms and associated benefits for disease diagnosis

Different AI algorithms have appreciable benefits in aiding in the diagnoses of RDs and non-RDs.[7] The ML helps in diagnosis via three types of algorithms: (a) unsupervised which works by identifying patterns, (b) supervised which classifies or predicts decisions based on former examples, and (c) reinforcement learning which uses reward and punishment processes to form a blueprint for operating in a definite obstacle. The first-generation AI works on making clinical decisions by analyzing large amounts of data for diagnosing the RDs. However, a key issue for RDs is a lack of large amounts of initial data which can be inputted. Whereas the second-generation AI works by looking for the clinical clues that are commonly overlooked in the early course of the identification of RDs. This algorithm efficiently helps in the early diagnosis for intervention and prevention of various RDs.[4] Detecting and aptly reading the diagnostic image is a necessity that is increasingly leading to the recognition of ML in AI for RDs.[1] ML, for example, has recently been utilized to identify which patients with systemic sclerosis are at a high risk of severe complications, early detection of organ involvement, and more.[3]

AI is an integral tool for the diagnosis of RDs as it can assist in image recognition, genetic analysis, and support clinical decision making. ML is a branch of AI that has demonstrated its effectiveness in systemic sclerosis diagnosis. ML is able to detect pulmonary involvement better than high-resolution computerized tomography (HRCT) and pulmonary function tests (PFTs), which currently serve as the standard diagnostic methods. Although HRCTs tend to be used when PFTs indicate a decline in respiratory function, ML algorithms may detect pulmonary involvement prior to the onset of deterioration resulting in improved survival rates and reduced health costs.[3] Deep learning (DL) represents a subset of ML algorithms involved in image recognition. DL algorithms break down complex mappings into clusters of simpler mappings to facilitate more efficient analysis. The algorithm can then use its visible layer to read the image and its hidden layers to extract important features from the image. By doing so, DL is able to provide superior recognition with extensive and high-dimensional data. In some cases, these images may contain important clinical features that can aid in rapid and accurate diagnosis. Compared with more traditional biometric methods, AI has shown greater flexibility and scalability which allows it to contribute to understanding complex relationships, improving early detection, and making routine tasks more efficient.[1]

Artificial intelligence systems utilizing genetic data

With respect to phenotypic and genetic analysis, several AI systems have shown their effectiveness in analyzing data to provide accurate diagnoses. As nearly 80% of RDs are genetic, AI has great potential in this arena.[5] Several tools have been applied to a variety of RDs. PhenIX, utilizes the Disease-Associated Genome, which combines phenotypic concepts with genetic information to diagnose Mendelian diseases. PhenIX effectively reads the patient's genetic sequence, identifies variants, and prioritizes them based on pathogenicity and similarity to the patient's phenotype. Similarly, the Xrare tool has been used to identify causative variants of Mendelian diseases by using similarity scores from phenotypic sets, genetic information from variant databases, and guidelines for variant prioritization. These tools can perform a variety of important functions, including predicting gene pathogenicity, discovery of molecular markers, and prediction models based on gene expression data.[7] These examples highlight the advantages of AI in quickly scanning large datasets and identifying variants of interest.[5]

There are other tools and applications, utilizing AI, that have been developed to aid in the initial diagnosis of RDs. Rare Disease Discovery and Genetic Disease Diagnosis based on Phenotypes are two such examples.[5] AI can also identify RD patterns utilizing Bayesian comorbidity networks. This type of AI methodology can be used to differentiate and group subtypes of both RDs and non-RDs. Bayesian comorbidity networks are also capable of identifying rare subtypes and comorbidity relationships. Bayesian information criteria can categorize the biological patterns in patients and cluster them accordingly. As a result, distinct subpopulation clusters can be identified which can contribute to a better understanding of the disease. Thus, AI not only aids in diagnosing rare diseases and their subtypes but also provides crucial insights for informed clinical decisionmaking.[7]

AI IN RARE DISEASE TREATMENTS

RDs have been neglected by the pharmaceutical industry due to their low and variable incidence, leaving

individuals with RDs often with few, expensive treatment options. However, based on previous studies and information from the National Institutes of Health (NIH) Office of Rare Diseases Research and the US Food and Drug Administration (FDA) Office of Orphan Products Development, it is clear that recent government support has led to the approval of many drugs for RDs, thanks in part to advances in AI technology.[8] Furthermore, the passage of the US Orphan Drug Act in 1983 and the European Union Regulation on Orphan Medicines in 2000 has rewarded innovation in rare disease treatment.[6] Despite this, the conventional drug discovery pipeline has significantly hindered the research and development of new drugs to treat RDs. The process is further considered challenging because of the potential for low revenue gains, but drug repurposing and the establishment of interdisciplinary centers for RDs have become trending topics to ameliorate this challenge.[8]

Artificial intelligence systems for rare disease treatment development

Two subtypes of AI, ML and DL, have proven useful in drug development. DL enables the creation of more tailored therapies, whereas ML is useful in clinical trials. AI has played a crucial role in a critical use case: forecasting patient susceptibility and identifying potential medications and therapies.[6] There are several DL algorithms that have been developed and applied to treatment development. Support Vector Machine is a type of DL algorithm that performs supervised learning and Random Forest makes output predictions by combining outcomes from a sequence of regression decision trees. These are the most commonly used AI approaches in rare and ultra-rare diseases, given their ability to handle complicated, high-dimensional data and images. These algorithms can also learn from limited datasets, which are often encountered in RDs due to their frequency. With the aid of AI-powered approaches, researchers are gaining a deeper understanding of the underlying causes of many illnesses, which aids in the identification of specific targets for therapy. Quantitative structure activity relationship (QSAR) modeling and high-throughput screening can generate large data sets, which AI can utilize for treatment development. Computational approaches, such as QSAR modeling, may also be utilized to generate novel treatment compounds with more desired properties. High-throughput screening campaigns can then generate large amounts of data, leading to the discovery of drugs such as riluzole for the treatment of amyotrophic lateral sclerosis.[8]

Advances in rare disease treatment development

The development of second-generation AI systems has enabled a patient-centered approach to the treatment and management of RDs. These systems aim to fill the gaps in diagnostic, prognostic, and therapeutic options by using a tailored closed-loop system to improve end-organ function and overcome tolerance or loss of effectiveness issues.[6] Some RDs do have proposed treatments or therapeutic approaches. However, due to the heterogeneity and complexity of RDs, many patients do not respond well to available therapies, including subsequent partial or complete loss of response.[4] Second-generation AI-based systems are increasingly patient-tailored allowing for a method to track and improve the outcomes of therapies.[4] Unlike previous approaches and iterations, these systems are not dependent upon large datasets and are dynamic systems that are able to adapt to ongoing changes in an RD patient's disease and therapy response.[4] Thus, these systems are particularly useful for tackling the issue of partial or complete loss of response by identifying these cases early.

Improvements to therapeutic and monitoring tools in RDs are critical. Second-generation AI tools, depending on the response to the therapeutic regimen, can finalize a treatment regimen along with electronic data and patient-reported outcomes. A calculation can adjust the treatment plan in reaction to changes in the patient's status, concurrent illnesses, drugs, natural conditions, and any other components that impact the patient's well-being and/or response to treatment.[4] The system helps patients at three levels: reminding patients of the dose and time of administration and incorporating non-pharmacological therapies, such as physiotherapy. It also includes a closed-loop system that adjusts the dose and delivery timing based on the patient's reaction to therapy and an algorithm that identifies disease-related patterns by determining variability in laboratory results or clinical parameters.[6] To underscore how critical monitoring of treatment regimens is for RDs, one must simply look to the example of Gaucher disease (GD). Current therapeutic options for GD include intravenous enzyme replacement therapy (ERT) and oral substrate reduction therapy, however, there is high interindividual variability in response to ERT.[4] This variability is associated with a high risk of long-term complications and thus it is critical to monitor the responses to these therapies systematically to discontinue therapy to avoid this high risk of long-term complications.[4]

Clinical trials are important for the development of treatments for RDs. Yet, traditional randomized clinical trials (RCTs) may not always be possible for individuals

with RDs. Clinical studies using AI have been used to overcome some of these limitations. Data mining computable phenotype algorithms have aided in the identification of patients with RDs for RCTs.[7] In the future, AI could also be utilized as a synthetic control, based upon data for expected disease progression, to overcome challenges with sufficient patient enrollment numbers. Furthermore, reliable biomarkers help in the identification of normal versus pathogenic processes and the assessment of the response to therapeutics or other interventions, which are both essential for the development of effective therapies.[7] A recent study of bone morphogenetic protein therapy for congenital pseudarthrosis of the tibia caused by neurofibromatosis type 1 used Ward hierarchical clustering to categorize subjects, and an RF-based algorithm identified possible biomarkers for therapeutic efficacy like the rate of cartilage formation (Pmc).[7] However, it is important to note that obtaining regulatory approval for an identified biomarker can often be a complex, lengthy process that must also be addressed. Overall, AI has the potential to play a role in the development of treatments for RDs, enabling a patient-centric approach that is tailored to the specific needs of each individual.

CONCLUSION

In this mini-review, we investigated AI and its potential for RD diagnosis and treatment. AI is rapidly growing in both efficiency and as a standard analytical tool for scientific discovery. Particularly for RDs, these advancements have great potential to create a more effective, efficient pipeline for RD studies, drug discovery, and therapeutic fine-tuning. AI has many use cases at all stages of development, including in discovery, preclinical, and clinical stages. In the discovery stage, AI may contribute to target identification, initial validation, and biomarker discovery. At the preclinical stage, it may be applied to test optimization, progression modeling, and safety profiles. Last, at the clinical stage, it can aid in patient recruitment, optimizing RCT, real-world data (RWD) analysis, diagnostic imaging, and developing precision medicine approaches.

Searching for published studies on AI and RDs supports the idea that a majority of studies on the topic have been published in just the past 3 years. Thirty-eight of the 45 studies (84%) published and indexed in PubMed on this particular topic were published from 2020 to 2023. Already DL has been utilized for epigenomic studies in RDs which could significantly boost discovery and therapy development.[9] Furthermore, quantitative model-based approaches like disease progression modeling with AI, advanced statistical approaches in natural history data, and RWD will play pivotal roles in increasing the efficiency of clinical

study design in the drug development process.[10] Already, there have been AI-based applications to optical coherence tomography, which demonstrate improvements in automated image pattern recognition.[11] Some clear assets of AI include natural language processing eliminating human variability in translating information from the medical record, assisting clinician decision-making processes for managing complications, and phenotypic prediction.[12–14]

Of the various types of AI that have emerged, each type of AI has its own strengths, limitations, and applications. Whereas DL is catered to the development of tailored therapies, ML is believed to be most applicable to clinical trials. Advances brought about by subsequent generations of AI are expected to more efficiently detect early signs of RDs, and test a plethora of candidates for drug development thereby maximizing the efficiency of the arduous process. It is important to note the limitations of some AI models. Namely, some AI models require accurate, clear, and large amounts of data in order to learn and perform best. The key challenge with RDs is that there is often not enough data available to effectively train a model. Many RDs have been difficult to treat because of the heterogeneity of their clinical presentation, treatment by different healthcare providers, and sometimes even their biological characteristics. It is imperative that more research on RDs is able to collect further data. It is important to note that different use cases and algorithms require a different amount of data to function at an optimal level. Recent advances have led to workarounds when data are limited, like data augmentation. Data augmentation allows the model to be trained and learn patterns from a smaller amount of data. Utilizing an unsupervised learning approach, patterns can even be identified from data that are unlabeled and complex. Another emerging approach is known as "fine-tuning" or transfer learning, wherein models can be trained on a big data set (such as a more common disease) and then retrained slightly using a smaller dataset (such as with RDs).[5] For example, the PLIER framework developed by researchers utilizes an unsupervised transfer learning framework and may be applied to smaller datasets, such as RDs and precision medicine.[15]

With a great deal of recent advances in AI model generation and application, it can be challenging to determine the best use case for each type of approach. Investigators often have to choose between complex, but high-accuracy AI models or simplified, less accurate models. On one hand, accuracy is crucial for effective early diagnosis and treatment, but if the outputs are too difficult to interpret then it can be hard to understand why a recommendation was made by the algorithm. However, future iterations of AI may improve accuracy and eventually become easier to train and utilize. In the future, AI is likely to contribute to earlier diagnosis of RDs which can improve prognosis, advance translational research, and help optimize precision

medicine. Already, AI has made significant strides in furthering our understanding of different diseases, fine-tuning therapeutic approaches, and identifying biomarkers that can be targets for new drug development.

FUNDING INFORMATION

CONFLICT OF INTEREST STATEMENT

The authors declared no competing interests for this work.

ORCID

Magda Wojtara https://orcid.org/0000-0001-6903-0206
Emaan Rana https://orcid.org/0009-0004-0483-4165
Taibia Rahman https://orcid.org/0009-0008-4951-800X
Palak Khanna https://orcid.org/0009-0008-9196-3318
Heshwin Singh https://orcid.org/0009-0000-1734-3593

REFERENCES

1. Huang B, Huang H, Zhang S, et al. Artificial intelligence in pancreatic cancer. *Theranostics*. 2022;12(16):6931-6954.
2. Krawitz PM. Artificial intelligence in the diagnosis of rare disorders: the development of phenotype analysis. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2022;65(11):1159-1163.
3. Bonomi F, Peretti S, Lepri G, et al. The use and utility of machine learning in achieving precision medicine in systemic sclerosis: a narrative review. *J Pers Med*. 2022;12(8):1198.
4. Hurvitz N, Azmanov H, Kesler A, Ilan Y. Establishing a second-generation artificial intelligence-based system for improving diagnosis, treatment, and monitoring of patients with rare diseases. *Eur J Hum Genet*. 2021;29(10):1485-1490.
5. Faviez C, Chen X, Garcelon N, et al. Diagnosis support systems for rare diseases: a scoping review. *Orphanet J Rare Dis*. 2020;15(1):94.
6. Visibelli A, Roncaglia B, Spiga O, Santucci A. The impact of artificial intelligence in the odyssey of rare diseases. *Biomedicine*. 2023;11(3):887.
7. Brasil S, Pascoal C, Francisco R, Dos Reis Ferreira V, Videira PA, Valadão AG. Artificial intelligence (AI) in rare diseases: is the future brighter? *Genes (Basel)*. 2019;10(12):978.
8. Alves VM, Korn D, Pervitsky V, et al. Knowledge-based approaches to drug discovery for rare diseases. *Drug Discov Today*. 2022;27(2):490-502.
9. Brasil S, Neves CJ, Rijoff T, et al. Artificial intelligence in epigenetic studies: shedding light on rare diseases. *Front Mol Biosci*. 2021;8:648012. doi:10.3389/fmolb.2021.648012
10. Liu J, Barrett JS, Leonardi ET, et al. Natural history and real-world data in rare diseases: applications, limitations, and future perspectives. *J Clin Pharmacol*. 2022;62(Suppl 2):S38-S55. doi:10.1002/jcph.2134
11. Petzold A, Albrecht P, Balcer L, et al. Artificial intelligence extension of the OSCAR-IB criteria. *Ann Clin Transl Neurol*. 2021;8(7):1528-1542. doi:10.1002/acn3.51320
12. James KN, Phadke S, Wong TC, Chowdhury S. Artificial intelligence in the genetic diagnosis of rare disease. *Clin Lab Med*. 2023;43(1):127-143. doi:10.1016/j.cll.2022.09.023
13. Ardahan Sevgili S, Şenol S. Prediction of chemotherapy-related complications in pediatric oncology patients: artificial intelligence and machine learning implementations. *Pediatr Res*. 2023;93(2):390-395. doi:10.1038/s41390-022-02356-6
14. Álvarez-Machancoses Ó, Fernández-Martínez JL. Using artificial intelligence methods to speed up drug discovery. *Expert Opin Drug Discov*. 2019;14(8):769-777. doi:10.1080/17460441.2019.1621284
15. Taroni JN, Grayson PC, Hu Q, et al. MultiPLIER: a transfer learning framework for transcriptomics reveals systemic features of rare disease. *Cell Syst*. 2019;8(5):380.e4-394.e4. doi:10.1016/j.cels.2019.04.003