DOI: 10.55522/jmpas.V14I6.6990 ISSN NO. 2320 – 7418

International peer reviewed open access journal

Journal of Medical Pharmaceutical and Allied Sciences

Journal homepage: www.jmpas.com CODEN: JMPACO

Review article

Artificial intelligence for orphan and rare diseases: a promising frontier in healthcare

Ajay Kumar Shukla*, Vimal Kumar Yadav

Institute of Pharmacy, Dr Rammanohar Lohia Avadh University, Uttar Pradesh, India

Corresponding author: Ajay Kumar Shukla, ✉ ashukla1007@gmail.com, **Orcid Id**: https://orcid.org/0000-0001-5689-3467

Institute of Pharmacy, Dr Rammanohar Lohia Avadh University Uttar Pradesh, India

Received - 20-11-2025**, Revised** - 28-11-2025, **Accepted** - 25-12-2025 **(DD-MM-YYYY)**

Refer This Article

Ajay Kumar Shukla, Vimal Kumar Yadav, 2025. Artificial intelligence for orphan and rare diseases: a promising frontier in healthcare. Journal of medical pharmaceutical and allied sciences, V 14 - I 6, Pages - 09 – 15. Doi: https://doi.org/10.55522/jmpas.V14I6.6990.

ABSTRACT

Due to their little knowledge and lack of available data, orphan and uncommon diseases which impact millions of people globally present particular difficulties in diagnosis, treatment, and research. Conventional methods in the medical field frequently fail to effectively handle these problems. But integrating artificial intelligence (AI) has great potential to change the way orphan and rare disease management is done. This study examines the state of the art and future directions for AI applications in the fight against chronic illnesses. Artificial intelligence (AI)-driven techniques, including computer vision, natural language processing, and machine learning, provide innovative solutions for managing rare and orphan diseases at different phases. They include using predictive analytics to detect diseases early, developing individualised treatment plans based on data-driven insights, and using computational techniques to speed up the development and repurposing of drugs. Furthermore, AI facilitates the combination and analysis of data from various sources, including real-world, clinical, imaging, and genomic data, enabling a more comprehensive understanding of these illnesses. Moreover, AI-driven platforms facilitate collaboration among researchers, physicians, patients, and industry participants, creating a mutually beneficial environment for information exchange, data sharing, and collective problem-solving. This paper demonstrates the concrete effects of AI in driving innovation in precision medicine, expediting the identification of uncommon diseases, and improving patient outcomes through case studies and examples. AI adoption in orphan and rare illnesses confronts a number of obstacles despite its enormous promise, including a lack of data, interpretability of algorithms, legal restrictions, and ethical issues. In order to ensure the ethical and fair integration of AI technology in healthcare, addressing these difficulties calls for interdisciplinary cooperation, strong governance structures, and continual stakeholder involvement. In conclusion, the management of orphan and uncommon diseases is being revolutionised by artificial intelligence, which presents hitherto unseen chances to improve diagnosis, therapy, and research results. Through the utilisation of AI-driven advancements, interested parties can work together to surmount the distinct obstacles presented by these illnesses and ultimately enhance the quality of life for millions of people afflicted with uncommon and orphan ailments.

Keywords: Artificial Intelligence, Orphan and Rare Diseases, disease diagnosis, artificial intelligence tool, impact of health disparities on emigration.

INTRODUCTION

This field has a lot of promise for artificial intelligence (AI), with a focus on deep learning. AI is already being successfully used in basic research, diagnosis, drug discovery, and clinical trials. AI technologies can be especially helpful for rare diseases (RDs), which are notably underrepresented in basic and clinical research. This field has a lot of promise for artificial intelligence (AI), with a focus on deep learning. AI is already being successfully used in basic research, diagnosis, drug discovery, and clinical trials. AI technologies are especially useful for rare diseases (RDs), which are notably underrepresented in basic and clinical research [1].

The realm of uncommon and ultra-rare diseases is where AI approaches are most frequently utilized such as Decision

trees, genetic boosting algorithms, and metric techniques, such as the K-nearest neighbor algorithm (k-NN), Support Vector Machines (SVM), statistical approaches, Bayesian models (BM), Artificial Neural Networks (ANN), and Ensemble methods are all examples of ML techniques, which have a wide range of applications across various disciplines are the most often used algorithms in this regard. These approaches are very popular because they can handle high-dimensional, complicated data; in fact, the most common type of input was discovered to be photographs. Large volumes of this kind of data are produced by medical imaging modalities as PET, CET, MRI, and ultrasound; as a result, AI systems can process them with ease because they are standardised. Additionally, these techniques can learn from small datasets, which is frequently the case with rare diseases because of their low prevalence. They can also recognise significant features that can aid in the better understanding of the disease's underlying mechanisms and possibly point to new treatment targets. The majority of research has concentrated on applying AI techniques for the diagnosis and prognosis of rare diseases, which is a common application of classification and prediction. This is also noteworthy, and it makes sense considering the difficulties in detecting and forecasting rare disorders. It is important to remember, though, that machine learning (ML) may also significantly contribute to bettering the course of treatment and hastening the development of new drugs by identifying promising candidates, estimating their efficacy, and determining the best dosages for them. It may even help find existing medications and repurpose them for the treatment of other illnesses [].

The scientific discipline of artificial intelligence (AI) seeks to build machines that can perceive and solve problems like humans. ML, in which computers learn from experience, and DL, which includes progressively more complicated models, are two of its subdivisions. In machine learning (ML), algorithms learn from data using semi-supervised, supervised, or reinforcement learning techniques. The process of machine learning comprises six essential elements: gathering data, preprocessing, standardisation, feature selection, model training, and assessment. Regression, grouping, and classification are examples of ML tasks. Applications for methods like SVM, genetic algorithms, and decision trees can be found in many different fields. Complex task addressing is made possible by DL, a subset of AI. Issues with data availability are impeding the adoption of AI in biology and biomedicine, despite its potential. Limited data make it difficult to identify therapeutic alternatives for rare disorders. Medic Kanren and other AI tools help connect genetic information to treatments. In order to treat ALS, IBM Watson analyses RNA-binding proteins. Gaucher disease and other uncommon diseases are predicted by GCAN. Compounds are screened for sialidosis, and cell-penetrating peptides are predicted for DMD treatment using Bayesian ML models and RF classifiers. Fanconi anaemia is predicted by multi-output regression machine learning. AKU patients' quality of life scores are predicted by an RF model based on their treatment regimen [3].

Rapid and efficient data collection, processing, and characterisation are necessary due to the increasing amount of biological data. This field has enormous promise for the application of artificial intelligence (AI), especially deep learning, which can help with fundamental research, diagnosis, drug discovery, and clinical trials. The use of AI technology has great potential to assist rare diseases (RDs), which are frequently overlooked in studies. Of the more than 7000 RDs worldwide, very few have access to therapies. AI's ability to combine and evaluate many data sources can help RDs overcome issues like dispersed patient populations and low diagnosis rates. This has the potential to completely change how RDs receive therapy. AI techniques are currently being used extensively in RD research, and this study seeks to summarise these developments. Included is a focus on congenital disorders of glycosylation (CDG), a subset of orphan RDs, highlighting its potential as a model for research on other RDs as well as common diseases [4].

The impact of health disparities on emigration is examined in this article, with a focus on families plagued by uncommon diseases for whom there are only limited treatment choices in certain nations. It seeks to arrange the currently available information on this rather unknown occurrence. Applying machine learning and text mining techniques to pertinent databases, the study finds patterns that are shared by patients, treatment reimbursement agencies, drug makers, and the larger context. Even while health-driven emigration isn't mentioned much directly, the authors clarify related terms and stress its importance in managing rare disease support systems, opening the door for more study in this area.

Artificial intelligence for orphan drug development

In order to illustrate the value of the recently developed quantitative retrospective natural history modelling (QUARNAM) approach, this research will examine a subset of 849 patients globally who suffer from seven uncommon neurogenetic illnesses. Comprehending the inherent course of uncommon illnesses is essential for customised therapies and family guidance. Comparable to a meta-analysis for individual studies, QUARNAM examines published case reports to answer important research issues like patient longevity, disease severity predictors, diagnostic time, and recruiting sources. It makes use of sophisticated statistical techniques such as regression analysis, cluster analysis, and Kaplan-Meier estimations. QUARNAM provides rapid insights into clinical outcomes with less effort than other approaches, like prospective studies or chart reviews, which makes it appropriate for drug development programs based on particular research needs [5].

We developed a novel approach to identify putative gene culprits in the genetically varied orphan illnesses, such as Emery-Dreifuss muscular dystrophy (EDMD), in an effort to tackle their complexity. The wide clinical spectrum of EDMD within families makes it challenging to treat. It is characterised by early onset contractures, increasing muscle atrophy, and heart conduction problems. Although the precise illness process is yet unknown, six known genes related to nuclear envelope proteins are associated to half of EDMD patients.

Our approach comprised building a primer library that targeted 301 genes in four categories: (I) genes linked to EDMD that are known to exist, (II) genes linked to similar muscular dystrophies, (III) candidates found in particular families through exome sequencing, and (IV) potential functional candidates, with a focus on muscle nuclear envelope proteins connected to the processes affected by EDMD. Then, using this library, 56 unrelated patients displaying symptoms similar to EDMD were sequenced [6].

Twenty-one patients were found to have distinguishable mutations, three of whom had previously unidentified mutations in recognised EDMD-related genes and eight of whom had mutations in genes connected to comparable muscular dystrophies. Moreover, a number of new candidate genes that our method discovered encode nuclear envelope proteins involved in gene regulation. Not only did our multifaceted method identify novel disease alleles, but it also illuminated other potential EDMD genes. Given the connection of identified candidates from the nuclear envelope to the plasma membrane, these results strongly imply that changes in gene regulation and mechanotransduction may be part of the pathomechanism of EDMD.

This approach shows potential for treating additional genetically heterogeneous orphan diseases and emphasises the importance of investigating related disorders using primer libraries. Support for this work was provided by the European Community's Seventh Framework Programme "Integrated European -omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NEUROMICS)" and the Wellcome Trust as well as Muscular Dystrophy UK and the Medical Research Council.

In biology, ontologies are essential for data annotation, integration, and analysis. They provide useful information about ontology classes and include both metadata in the form of annotation axioms and formally structured axioms. Typically, these annotation axioms consist of synonyms, descriptions, and class labels. Semantic similarity measures and other ontology-based analytic techniques frequently underutilise this metadata despite its rich semantic richness [7].

We offer a new technique for creating vector representations of biological elements inside ontologies: OPA2Vec. Formal ontology axioms are combined with annotation axioms from the ontology metadata in OPA2Vec. We generate feature vectors by using a pre-trained Word2Vec model on either a corpus or abstracts/full-text articles. We employ two methods to verify OPA2Vec:

Protein-protein interactions are predicted using similarity metrics and protein vector representations from two datasets. Applying phenotypic similarity to forecast associations are between genes and diseases. We generate vector representations of genes and diseases using a phenotypic ontology, and we infer links between them based on the phenotypes of mouse models. In terms of predicting gene-disease connections, OPA2Vec performs noticeably better than current techniques. We use OPA2Vec to find potential genes for a wide range of uncommon and orphan diseases by utilising data from mouse models. Of particular note is OPA2Vec's flexibility in creating vector representations for any biological entity with any biomedical ontology.

Identifying rare diseases

A retrospective analysis was conducted to identify possible Pompe disease patients by screening 350,116 electronic health records (EHRs). By applying Symptoma's AI to anonymised EHRs from the University Hospital Salzburg clinic group, 104 patients were identified as potentially having Pompe disease. Nineteen cases were found to be clinically probable for Pompe illness after manual examination by clinicians; this led to an AI specificity of 18.27%. One in every 18,427 persons in the greater Salzburg area was thought to have Pompe disease. Based on when symptoms first appeared, patient cohorts were divided into groups that had either infantile-onset (IOPD) or late-onset (LOPD) Pompe disease. This study demonstrates how well Symptoma's AI can identify individuals with rare diseases using electronic health records (EHRs); on average, 5.47 patients are screened for every suspected candidate.

The National Institutes of Health (NIH) established the Genetic and Rare Diseases (GARD) Information Centre, which provides extensive health information about more than 6500 rare and genetic diseases. Traditional approaches to gathering and organising scientific data, however, are unable to keep up with the rapid advancement of science. We have investigated computational methods to improve information curation to address this. Our goal is to provide a foundational database for translational science researchers and enhance the sustainability and relevance of consumer health information [8].

We have used Neo4j to create a meta-ontology-based knowledge network for uncommon diseases that incorporates information from 34 biomedical sources. More than 3.8 million nodes and 84 million relations, including information on rare diseases and

carefully selected pharmacological correlations, are present in this graph. We have connected UNII codes to FDA-designated medications and GARD disorders to FDA orphan designations through semi-automated mappings. We've also included drug indication connections from Inxight Drugs.

We have illustrated the graph's potential to expedite the curation of scientific knowledge on rare diseases through four case studies by creating disease mappings/profiles and discovering pathogenesis linkages. Our integrated knowledge graph, which includes a large amount of research and biomedical data, demonstrates how graph-based resources can be used to help consumer health information providers such as GARD. Our method emphasises immediate applications, but it also shows how difficult it is to sustain knowledge when the body of information is constantly changing. Our effective integration of datasets into an openly accessible knowledge graph has created new avenues for rare illness translational science research. Today, scientists can recognise the traits of diseases that are essential for applying findings in various study fields [9].

Diagnosis

The unusual symptoms and scarcity of clinical data associated with rare diseases make diagnosis difficult. Analysing big datasets to find patterns in illness states is one area where AI, especially machine learning techniques, shows potential. Accurate diagnosis can be aided even with minimal training data by methods such as collaborative filtering with phenotypic information and machine learning models. Potential disease markers for Huntington's disease (HD) have been identified through the use of machine learning techniques such as SVM and DL for gait assessment and oculomotor function analysis. Applying SVM and CNN algorithms, comparable methods have been used to distinguish multiple system atrophy (MSA) from Parkinson's disease (PD).

Deep CNNs and RF models have been used to classify neuroimaging data in ALS. SVM techniques have been applied extensively to improve diagnostic accuracy based on proteomics or particular biomarkers in a number of uncommon disorders, such as osteogenesis imperfecta, Friedreich ataxia, amyloidosis, and hypophosphatasia. Because ANN and DL models have excellent sensitivity and specificity, they are being used more and more for illness categorization utilizing several imaging modalities, like MRI and eye photos. When applied to rare illness scenarios, ensemble learning techniques improve diagnostic accuracy by pooling the predictions of numerous models. Examples include discovering microbiological biomarkers for primary sclerosing cholangitis (PSC) and using EL to select miRNAs for pulmonary arterial hypertension (PAH). ML methods like RF also aid in early detection of photoreceptor integrity deterioration in retinal diseases and classification of rare myopathic conditions

The most important factor influencing time and cost in genome-based diagnosis of uncommon genetic illnesses is now the clinical interpretation of genetic variants in connection to a patient's phenotype. By fusing growing knowledge about genetic diseases with prediction techniques, artificial intelligence (AI) presents a promising way to streamline and expedite genome interpretation. In this work, we assessed Fabric GEM's diagnostic performance. Fabric GEM is a novel AI-powered clinical decision support tool that aims to speed up genomic interpretation. Our evaluation was based on a retrospective cohort of 119 probands who underwent whole-genome or whole-exome sequencing (WGS, WES) and were mostly NICU newborns with uncommon genetic illness diagnoses. We further confirmed our results with 60 additional cases from five academic medical centres in a different cohort. We examined trio, duo, and singleton scenarios for comparison, utilising the most advanced variant prioritising techniques available today. The variants that resulted in diagnoses included structural variants (SVs) and varied inheritance mechanisms. Phenotypes of patients were either manually or automatically retrieved from clinical notes using clinical natural language processing (CNLP). Moreover, we reexamined 14 cases that remained unresolved [10].

Fabric GEM prioritised a median of three candidate genes per instance, regardless of phenotype descriptions acquired manually or by CNLP, and consistently rated over 90% of causative genes among the top or second possibilities. Analysis as a singleton revealed that it retained the same ranking for trios and duos. Whether or not GEM provided or inferred the causal SVs, in cases having diagnostic SVs, the causal SVs were identified by GEM as the best candidate in 17 out of 20 cases and within the top five in 19 out of 20 cases. Even without the genotypes of the parents, it performed similarly. Reanalysis of previously unsolved cases led to a novel finding in one case, no further advancements in three cases upon manual review, and no new findings in 10 cases.

Fabric GEM automatically nominated a short selection of candidate genes and illnesses for final review and reporting, facilitating diagnostic interpretation across all variation types. When combined with CNLP's deep phenotyping, it offers a great deal of automation potential for diagnosing genetic diseases, which might save expenses and speed up case evaluation [11].

Most recently, we have made progress in understanding, characterizing, and treating this illness. chronic thromboembolic pulmonary hypertension (CTEPH) imaging features have been made easier to identify by guidelines that support systematic surveillance post-acute pulmonary embolism and the understanding that CTEPH can first present as acute pulmonary embolism. Artificial intelligence in diagnosis has made it possible to diagnose diseases early and more quickly. This Series paper shows how the realization

that CTEPH is a result of inflammatory thrombosis has led to the development of efficient multidisciplinary care that combines medicinal, surgical, and interventional methods. We provide imaging samples that emphasise important vascular targets, describe microvascular targets, describe the instruments that are available, and provide an algorithm for a first treatment plan that was developed through interdisciplinary talks at CTEPH centres. Additional investigation is required to improve the application and combination of multimodal therapy in CTEPH to improve long-term patient outcomes [12].

Prognosis

Predicting the course, duration, and outcome of a rare disease requires expertise in prognostication; treatment is frequently prioritised before cure. AI, and ML in particular, helps in prognosis by finding patterns and making predictions based on the analysis of large data sets, such as genomic and electronic health record data. Current research on alkaptonuria (AKU) and adrenocortical cancer (ACC) demonstrates the potential of machine learning (ML) in locating biomarkers and enhancing prognostic precision. While research on Huntington's disease (HD) uses SVMs and hybrid models for prognostication, studies on ALS use AI to estimate survival times and evaluate the severity of the condition. SVMs and boosting algorithms have shown useful in a number of situations, including the prediction of energy expenditure in patients with Duchenne muscular dystrophy (DMD) and the risk of hepatic decompensation in patients with primary sclerosing cholangitis (PSC). These illustrations show the variety of uses of AI in rare disease prognosis [13].

Challenges of AI

All age groups are affected by rare diseases, which pose serious medical, social, and financial difficulties for sufferers, their families, and caretakers. There are still a few viable treatments for the about 7000 uncommon diseases, despite efforts to improve diagnostic skills. A disease is considered rare in the United States if it affects fewer than 200,000 people, whereas it is considered rare in the European Union if it affects fewer than 5 people out of 10,000. Remarkably, 90% of uncommon diseases have no viable therapy. Raising awareness, advocating for, and reaching out to marginalised people around the world—such as those with poor incomes, low literacy rates, affiliations to minority ethnic groups, living in both urban and rural areas, as well as in emerging countries is vital. Regulatory actions and increased research efforts are complemented by treating patients as co-investigators in the process, which speeds up the evaluation and approval process for medications that treat critical or life-threatening disorders. Untreated patients have hope thanks to the developing pipeline of novel treatments. Advancements in artificial intelligence, machine learning, and medical bioinformatics, along with the availability of large datasets, lead to the discovery of novel treatment candidates for screening and assessment. Advanced analytics support the creation of research hypotheses, care guidelines, therapy response evaluation, illness pattern recognition, and progression prediction within particular patient populations [14].

All age groups are affected by rare diseases, which pose serious medical, social, and financial difficulties for sufferers, their families, and caretakers. There are still a few viable treatments for the about 7000 uncommon diseases, despite efforts to improve diagnostic skills. A disease is considered rare in the United States if it affects fewer than 200,000 people, whereas it is considered rare in the European Union if it affects fewer than 5 people out of 10,000. Remarkably, 90% of uncommon diseases have no viable therapy. Raising awareness, advocating for, and reaching out to marginalised people around the world—such as those with poor incomes, low literacy rates, affiliations to minority ethnic groups, living in both urban and rural areas, as well as in emerging countries is vital. Regulatory actions and increased research efforts are complemented by treating patients as co-investigators in the process, which speeds up the evaluation and approval process for medications that treat critical or life-threatening disorders. Untreated patients have hope thanks to the developing pipeline of novel treatments. Advancements in artificial intelligence, machine learning, and medical bioinformatics, along with the availability of large datasets, lead to the discovery of novel treatment candidates for screening and assessment. Advanced analytics support the creation of research hypotheses, care guidelines, therapy response evaluation, illness pattern recognition, and progression prediction within particular patient populations [15].

There are many obstacles in the way of rare disease research. The need for innovation is evident, given that the development of orphan drugs sometimes takes more than 15 years, from the initial patent application to the debut of the medicine, and that diagnosis can take decades. For a long time, patient advocacy groups (PAGs) have been crucial in assisting researchers in establishing communication with the rare disease population. But although PAGs play such an important role, working together with researchers can still be difficult, which emphasises the need for innovative approaches to build trust and expedite procedures. One potential option is the use of AI-driven clinical solutions, which provide automated workflows and omnichannel communication to improve researcher-PAG collaboration. These tools offer to address the logistical and representational issues that the profession is facing in addition to facilitating improved cooperation. the significance of using AI to address these problems, especially in improving inclusion and ethnic representation in clinical trials. As he noted, there is an indisputable need for a more accessible and egalitarian strategy, given that only

25% of participants were non-White and that half of the patients with uncommon diseases were youngsters.

Unstructured and poorly standardised textual data, such as medical records, is a difficulty for machine learning models and can cause overfitting. Because biological mechanisms of diseases are complex, only a few algorithms investigate them. Developers need to design procedures that tackle these issues and guarantee dependability by offering justifications and carrying out thorough error analysis. Their usefulness in clinical practice is further hampered by the paucity of studies validating their models on external datasets. It takes cooperation between data scientists, researchers, and physicians to create standardised procedures and support external validation studies [16].

Clinical neurophysiology AI approach

In the past, electrophysiological methods were the mainstay of sleep medicine. However, current trends suggest that behavioural and self-reported metrics are becoming more important, and neurophysiological research is being neglected. This study examines 18,370 PubMed-indexed publications on sleepiness in an attempt to verify this notion. We were able to identify notions about tiredness and the methods of study by using Natural Language Processing. The decline of neurophysiological techniques in favor of behavioural assessments is confirmed by the results. Two approaches are suggested to incorporate neurophysiology into sleep medicine: comparing clinical neurophysiology with dimensional sleepiness models and reanalyzing electrophysiological data from routine testing [17].

AI in drug repositioning of drug

One essential way to meet the unmet therapeutic requirements in rare diseases is through repurposing. We have highlighted important lessons by looking at previous cases of repurposing medications for limited populations and related studies. These include the significance of encouraging public-private collaborations, obtaining early regulatory advice, and exchanging clinical insights. We've also evaluated the state of drug repurposing for uncommon diseases, emphasizing the benefits of artificial intelligence, real-world data use, innovation platforms, patient involvement, and regulatory measures. These developing strategies have the potential to improve pharmaceutical development, public health, patient care, and drug repurposin. ongenital diseases of glycosylation (CDG) are a group of uncommon genetic abnormalities affecting glycosylphosphatidylinositol anchor production and protein and lipid glycosylation. Recent advances in research have greatly aided in the discovery of therapies for these disorders. The practice of repurposing current medications for innovative medical uses, or drug repositioning, is becoming more popular for both common and uncommon illnesses. Using artificial intelligence (AI), a systematic search for repositioned molecules is the newest breakthrough. Drug repositioning speeds up drug discovery and lowers costs when compared to standard methods, which is beneficial for uncommon disorders like CDG in particular. AI systems have shown promise in the diagnosis, categorisation, and characterisation of diseases, as well as in the eventual discovery of treatments for uncommon ailments. For the purpose of improving therapeutic development, biomarker availability and reliable disease models are essential, especially in rare and heterogeneous disorders like CDG. This study explores the literature on medications that have been repositioned for CDG, whether they were found systematically or by chance. It also summarises current developments in disease models, biomarkers, and stakeholder viewpoints about AI's potential to aid in the development of CDG therapies [18].

CONCLUSION

ML techniques analyse enormous databases and electronic health information to help identify and diagnose uncommon diseases. Much with DL, predictive modelling predicts the course of an illness, allowing for tailored therapies and early interventions. With AI, patient subpopulations are found for clinical trials that are more effective. Annotated, limited data sets and interpretable models present challenges. To enhance modelling efforts, open data and patient registries can be integrated. The understanding of patients and clinicians depends on explainable AI. Bias avoidance, ethical application, and transparency are key to ML's potential in rare diseases. Safety and dependability are ensured via validation via trials and regulatory frameworks. For real-world patient benefits, overcoming data obstacles and maintaining fairness are crucial.

REFERENCES

1. Bishop C M, 2006. Pattern Recognition and Machine Learning. Information Science and Statistics. Pages 31073-2

2. Brasil S, Allocca M, Magrinho SCM, et al, 2022. Systematic Review: Drug Repositioning for Congenital Disorders of Glycosylation (CDG). Int J Mol Sci. 23(15), Pages 8725. Doi: 10.3390/ijms23158725.

3. Brasil S, Pascoal C, Francisco R, et al, 2019. Artificial Intelligence (AI) in Rare Diseases: Is the Future Brighter? Genes (Basel). 10(12), Pages 978. Doi: 10.3390/genes10120978.

4. De La Vega FM, Chowdhury S, Moore B, et al, 2021. Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. Genome Med. 13(1), Pages 153. Doi: 10.1186/s13073-021-00965-0.

5. Delcroix M, de Perrot M, Jaïs X, et al, 2023. Chronic thromboembolic pulmonary hypertension: realising the potential of multimodal management. Lancet Respir Med. 11(9), Pages 836-850. Doi: 10.1016/S2213-2600(23)00292-8.

6. Garbade SF, Zielonka M, Komatsuzaki S, et al, 2021. Quantitative retrospective natural history modeling for orphan

drug development. J Inherit Metab Dis. 44(1), Pages 99-109. Doi: 10.1002/jimd.12304.

7. Groft SC, Posada M, Taruscio D, 2021. Progress, challenges and global approaches to rare diseases. Acta Paediatr. 110(10), Pages 2711-2716. Doi: 10.1111/apa.15974.

8. Jonker AH, O'Connor D, Cavaller-Bellaubi M, et al, 2024. Drug repurposing for rare: progress and opportunities for the rare disease community. Front Med (Lausanne). 11, Pages 1352803. Doi: 10.3389/fmed.2024.1352803.

9. Künnemann K, 2019. 10th World Orphan Drug Congress (WODC) (November 12-14, 2019 - Barcelona, Spain). Drugs Today (Barc). 55(12), Pages 753-758. Doi: 10.1358/dot.2019.55.12.3107705.

10. Lapidus D, 2023. Strengths and limitations of new artificial intelligence tool for rare disease epidemiology. J Transl Med. 21(1), Pages 292. Doi: 10.1186/s12967-023-04152-0.

11. Lin S, Nateqi J, Weingartner-Ortner R, et al, 2023. An artificial intelligence-based approach for identifying rare disease patients using retrospective electronic health records applied for Pompe disease. Front Neurol. 14, Pages 1108222. Doi: 10.3389/fneur.2023.1108222. eCollection 2023.

12. Martin VP, Gauld C, Taillard J, et al, 2024. Sleepiness should be reinvestigated through the lens of clinical neurophysiology: A mixed expertal and big-data Natural Language Processing approach. Neurophysiol Clin. 54(2), Pages 102937. Doi: 10.1016/j.neucli.2023.102937.

13. Meinke P, Kerr ARW, Czapiewski R, et al, 2020. A multistage sequencing strategy pinpoints novel candidate alleles for Emery-Dreifuss muscular dystrophy and supports gene misregulation as its pathomechanism. EBioMedicine. Pages 102587. Doi: 10.1016/j.ebiom.2019.11.048.

14. Schreidah CM, Reynolds GB, Fahmy LM, et al, 2023. Scoping review of genetic databases for rare dermatologic diseases: Opportunity for artificial intelligence and machine learning. JAAD Int. 12, Pages 24-31. Doi: 10.1016/j.jdin.2023.02.017.

15. Skweres-Kuchta M, Czerska I, Szaruga E, 2023. Literature Review on Health Emigration in Rare Diseases-A Machine Learning Perspective. Int J Environ Res Public Health. 20(3), Pages 2483. Doi: 10.3390/ijerph20032483.

16. Smaili FZ, Gao X, Hoehndorf R, 2019. OPA2Vec: combining formal and informal content of biomedical ontologies to improve similarity-based prediction. Bioinformatics. 35(12), Pages 2133-2140. Doi: 10.1093/bioinformatics/bty933.

17. Toh T S, Dondelinger F, Wang D, 2019. Looking beyond the hype: Applied AI and machine learning in translational medicine. EBioMedicine Pages 607–615.

18. Zhu Q, Nguyen DT, Grishagin I, 2020. An integrative knowledge graph for rare diseases, derived from the Genetic and Rare Diseases Information Center (GARD). J Biomed Semantics. 11(1), Pages 13. Doi: 10.1186/s13326-020-00232-y.