International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Navigating the uncharted territory of rare diseases: leveraging artificial intelligence to achieve groundbreaking advances, with a focus on Congenital Disorders of Glycosylation (CDG): A comprehensive analysis

Vibhuti Choubisa * and Karuna Soni

Department of Computer Science and Engineering, Pacific Academy of Higher Education and Research University, Udaipur, India.

International Journal of Science and Research Archive, 2024, 12(02), 1061–1071

Publication history: Received on 31 May 2024; revised on 08 July 2024; accepted on 10 July 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.12.2.1255

Abstract

Artificial Intelligence (AI) is changing the game when it comes to researching, diagnosing, and treating rare genetic disorders, bringing new hope for better patient outcomes. This paper focuses on how AI is making a big difference in tackling Congenital Disorders of Glycosylation (CDG), a group of over 130 rare genetic disorders that disrupt glycosylation, a critical biological process. By using advanced machine learning algorithms, AI tools like PredictSNP, REVEL, and Face2Gene are improving the accuracy of diagnosing these disorders, making it quicker and easier to identify them early. We explore how AI helps predict glycosylation sites, identify important Golgi proteins, and classify different disease phenotypes, offering new insights into how these diseases work. Additionally, AI is playing a crucial role in finding new treatments, including repurposing existing drugs to treat CDG. Despite the exciting progress, there are still challenges to overcome, such as ensuring high-quality data, making AI models understandable, and addressing ethical concerns. To truly unlock the potential of AI in this field, we need to integrate data from various sources and establish strong ethical guidelines. This paper highlights the importance of collaboration among researchers, clinicians, and policymakers to fully leverage AI's capabilities, paving the way for innovative and effective solutions for managing CDG and other rare genetic disorders.

Keywords: Artificial intelligence; Big data; Congenital disorders of glycosylation; Diagnosis; Drug repurposing; Machine learning; Personalized medicine; Rare diseases

1. Introduction

The landscape of medical diagnostics and therapeutics is undergoing a profound transformation, driven by advancements in artificial intelligence (AI) and machine learning. This revolution is particularly impactful in the realm of rare genetic disorders, where traditional methods often fall short due to the complexity and rarity of these conditions. Among the myriads of rare genetic diseases, Congenital Disorders of Glycosylation (CDG) stand out due to their intricate genetic and phenotypic variability. CDG represents a group of over 130 genetic disorders characterized by defects in glycan synthesis and attachment, leading to widespread and varied clinical manifestations. The diagnosis and treatment of CDG are fraught with challenges, primarily due to the diverse genetic mutations and the subtle, often overlapping symptoms that characterize these disorders.

In recent years, AI has emerged as a pivotal tool in addressing these challenges. By leveraging machine learning algorithms and vast datasets, AI enables more accurate and efficient assessment of genetic variant pathogenicity, thereby enhancing diagnostic precision and therapeutic strategies. Tools such as PredictSNP, REVEL, and VEST utilize advanced computational methods to predict the pathogenic potential of genetic variants, providing critical insights into

* Corresponding author: Vibhuti Choubisa

disease mechanisms and aiding in the formulation of personalized treatment plans. Moreover, specialized AI applications like Face2Gene harness facial recognition technology to identify characteristic features associated with specific genetic disorders, including CDG, further streamlining the diagnostic process.

This paper delves into the utilization of AI in navigating the frontiers of rare diseases, with a particular focus on CDG. It explores the role of AI in assessing genetic variant pathogenicity, highlights the capabilities and limitations of current AI tools, and presents a case study on the application of AI in diagnosing and managing CDG. By examining these aspects, we aim to underscore the transformative potential of AI in improving the outcomes for patients with rare genetic disorders, ultimately advancing the field of precision medicine.

2. Understanding CDG

2.1. What is CDG

Congenital Disorders of Glycosylation (CDG) are a group of over 130 rare genetic disorders that affect how sugars are attached to proteins and fats in the body. This process, called glycosylation, is crucial for the proper function of many proteins and fats.[1] When glycosylation doesn't work correctly, it can lead to a variety of health problems that affect many different parts of the body. People with CDG can experience a wide range of symptoms, including developmental delays, neurological issues, liver problems, blood clotting disorders, and immune system deficiencies.

2.2. Genetic Basis and Variability

CDG is caused by mutations in genes that are responsible for glycosylation. These genetic changes disrupt the normal pathways for attaching sugars to proteins and fats, leading to defective glycosylation. Because there are many different genes involved in glycosylation, the mutations can vary widely, resulting in a broad spectrum of symptoms and disease severity among patients.[2] This genetic and phenotypic variability makes CDG particularly challenging to diagnose and treat.

2.3. Clinical Manifestations

The symptoms of CDG can be very diverse and range from mild to severe. Common issues include:

- Neurological Problems: Developmental delays, intellectual disabilities, seizures, and low muscle tone.[3]
- Growth and Development Issues: Poor growth, feeding difficulties, and delayed motor skills.
- Liver Problems: Enlarged liver, abnormal liver enzymes, and blood clotting disorders.
- Blood Clotting Issues: Unusual bleeding and clotting problems.
- Immune System Issues: Frequent infections and immune system abnormalities.

Because the symptoms of CDG overlap with many other conditions, it can often be misdiagnosed or go undiagnosed for a long time.[4]

2.4. Diagnostic Challenges

Diagnosing CDG is difficult due to the wide range of symptoms and the rarity of the disorder. Traditional methods include biochemical tests to detect abnormal glycoproteins and genetic tests to identify mutations. However, these methods can be expensive, invasive, and time-consuming. Additionally, because CDG is so rare, many doctors may not be familiar with it, making diagnosis even more challenging.

2.5. Current Treatment Approaches

There is currently no cure for CDG, so treatment focuses on managing symptoms and improving the quality of life for patients. Treatment options include:

- Supportive Care: Physical therapy, occupational therapy, and nutritional support to help with developmental and growth issues.
- Medical Management: Medications to control seizures, manage liver problems, and address blood clotting disorders.
- Gene Therapy: Experimental approaches are being explored to try and correct the underlying genetic defects.[5]

Ongoing research aims to better understand CDG and develop more effective treatments. As discussed in the following sections of this paper, AI has the potential to significantly advance these efforts, offering new hope for patients and their families.

3. AI in Rare Disease Research

3.1. Literature Review on AI in Rare Diseases

Table 1 AI and ML-Based Methods in Rare Diseases

Function	Specific Function	Software/Platform/Algorithm	AI/ML Method	Disease(s)	Reference
Mutation Detection and Prediction	Predicts pathogenicity of genetic variants	CNVdigest, DNorm, Fathmm-MKL, M-CAP, MetaLR, Meta-SVM, PredictSNP, SpliceAI, VAAST Variant Prioritizer (VVP)	Various ML algorithms	Digeorge syndrome, Mevalonic kinase deficiency, RDs with intellectual disability and autism spectrum disorders	Yan et al., Alirezaie et al., Browne et al., Jaganathan et al., Alirezaie et al.
Phenotype-Driven Diagnosis	HANRD system	Ontological and curated associations	Information propagation algorithm	Various RDs	[35]
	Rule mining for skeletal dysplasia	Mixed algorithm	Rule mining, Dempster–Shafer theory	Skeletal dysplasia	[36]
	Naïve Bayes classifier for MPS II	Naïve Bayes classifier	Literature data training	Mucopolysaccharidosis type II	[37]
	DDSS with NLP	Collaborative filtering, NLP	Information extraction	Various RDs	[38, 39]
Imaging-Based DDSS	DeepGestalt for facial analysis	Deep learning framework	Face recognition transfer learning	Various genetic syndromes	[45]
	SVMs for acromegaly	SVMs	Facial image analysis	Acromegaly	[51]
Biochemical Fingerprinting	IR spectroscopy for HHT	Artificial neural networks	IR spectroscopy	Hereditary hemorrhagic telangiectasia	[54]
Prognostic Markers	Deep learning for synovial sarcoma	Deep learning	Survival prediction algorithm	Synovial sarcoma	[55]
Disease Mechanisms	MultiPLIER for biological pathways	Deep learning	Gene expression analysis	Various RDs	[53]
Disease Categorization	CFML for bone dysplasia	Characteristic feature mining	ML-based rule mining	Bone dysplasia disorders	[54]

3.2. AI in Rare Disease Treatment

Several studies have highlighted the potential of AI in the diagnosis and management of rare diseases. Behrens and Koretzky (2017) discussed the concept of precision medicine in the context of cytokine storm syndrome, emphasizing the need for personalized and targeted approaches to treatment. Similarly, Baumbusch, Mayer, and Sloan-Yip (2018) examined the experiences of parents navigating the healthcare system for children with rare diseases, underscoring the need for innovative solutions to improve patient care.[6,7]

Furthermore, Porumb et al. (2020) conducted a pilot study on the use of deep learning for the detection of hypoglycemic events based on electrocardiogram (ECG) data, demonstrating the potential for AI to aid in the early detection of rare disease-related complications. These findings suggest that AI has the capacity to enhance diagnostic accuracy, predict disease progression, and optimize treatment strategies for individuals with rare diseases.[8]

3.2.1. AI-Enabled Precision Medicine and Multidisciplinary Management

In the realm of rare disease management, Scarpa et al. (2011) provided European recommendations for the diagnosis and multidisciplinary management of mucopolysaccharidosis type II, highlighting the importance of a multidisciplinary approach to care. The integration of AI into precision medicine has the potential to streamline multidisciplinary collaboration, facilitate data-driven decision-making, and improve patient outcomes for individuals with rare diseases (Behrens & Koretzky, 2017).[9,10]

3.2.2. AI in Rare Disease Diagnosis and Prediction

Agrahary (2020) explored the use of machine learning algorithms for heart disease prediction, shedding light on the potential of AI to predict disease onset, progression, and response to treatment.[11] Additionally, Huynh et al. (2020) delved into the application of AI in radiation oncology, demonstrating the role of AI in personalized treatment planning and delivery. These insights underscore the potential of AI to revolutionize rare disease diagnosis, prediction, and treatment across various medical specialties.[12]

3.2.3. Knowledge Gaps

While the existing literature provides valuable insights into the application of AI in rare disease treatment, several knowledge gaps exist. Future research should focus on the development of AI-driven diagnostic tools for rare diseases that integrate multimodal data sources, including genetic, clinical, and imaging data. Additionally, the implementation of AI-enabled predictive models for rare disease progression and treatment response warrants further investigation.[13] Moreover, the ethical, legal, and social implications of AI in rare disease management remain understudied and require attention in future research endeavors. The integration of AI into rare disease treatment holds immense promise for improving diagnostic accuracy, personalizing treatment approaches, and optimizing patient care. [14] However, further research is needed to address knowledge gaps and propel the field of AI in rare disease treatment forward.[63,64]

Overall, this literature review has highlighted the potential of AI to transform the landscape of rare disease diagnosis, management, and treatment. By synthesizing and integrating various research findings, this review has shed light on the current state of AI in rare disease treatment, identified knowledge gaps, and proposed future research directions. As AI continues to evolve, its impact on rare disease care is expected to grow, ultimately leading to improved outcomes for individuals with rare diseases.[15]

3.2.4. Mutation Detection and Prediction

AI tools like CNVdigest, PredictSNP, and SpliceAI employ various ML algorithms to predict the pathogenicity of genetic variants across different RDs. These tools integrate data from diverse sources to enhance diagnostic accuracy and facilitate personalized treatment strategies.

3.2.5. Phenotype-Driven Diagnosis

Phenotype-driven systems such as HANRD and DDSS utilizing NLP and collaborative filtering improve disease prioritization and diagnostic accuracy by integrating clinical and genetic data. AI-based tools like DeepGestalt analyze facial images to identify genetic syndromes, enhancing clinical phenotype recognition.[16,17,18]

3.2.6. Imaging-Based DDSS

AI algorithms, including SVMs and deep learning frameworks like DeepGestalt, analyze medical images to assist in RD diagnosis. These tools recognize characteristic features associated with genetic syndromes, aiding clinicians in accurate disease identification.[20,21]

3.2.7. Biochemical Fingerprinting

AI applications like IR spectroscopy combined with neural networks offer alternative diagnostic methods for diseases like hereditary hemorrhagic telangiectasia (HHT), where genetic testing may be inconclusive or costly.[22]

3.2.8. Prognostic Markers

Deep learning algorithms predict disease progression and therapeutic responses in RDs, guiding personalized treatment decisions and improving patient outcomes.

3.2.9. Disease Mechanisms and Categorization

AI-driven approaches elucidate disease mechanisms through gene expression analysis (e.g., MultiPLIER) and automate disease categorization using ML-based rule mining (e.g., CFML), thereby advancing our understanding of complex RDs and facilitating targeted therapies.[23,24]

3.3. AI Techniques Applicable to Congenital Disorders of Glycosylation (CDG)

Congenital Disorders of Glycosylation (CDG) constitute a heterogeneous group of metabolic disorders characterized by defects in glycosylation processes, impacting various physiological systems.[25,26] The application of Artificial Intelligence (AI) techniques offers promising avenues for enhancing understanding, diagnosis, and therapeutic strategies for CDG. This review discusses key AI methodologies pertinent to CDG research and clinical practice.[27]

3.3.1. Genetic Variant Analysis

AI-driven machine learning algorithms, such as Random Forests, Support Vector Machines (SVMs), and Neural Networks, play pivotal roles in analyzing genomic data associated with CDG.[28] These algorithms excel in identifying pathogenic variants within the complex genomic landscape of CDG, enabling precise genotype-phenotype correlations crucial for disease characterization and personalized treatment strategies.

3.3.2. Phenotype-Genotype Integration

Integration of phenotypic data with genomic information using AI-based ontologies and knowledge graphs facilitates comprehensive understanding of CDG.[29,30] By leveraging structured ontological frameworks, AI enhances the discovery of intricate genotype-phenotype relationships, thereby refining diagnostic accuracy and guiding therapeutic interventions tailored to individual patient profiles.

3.3.3. Protein Structure Prediction and Analysis

Deep Learning models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are instrumental in predicting and analyzing protein structures affected by CDG-associated mutations. These AI techniques elucidate the structural and functional consequences of glycosylation-related protein alterations, offering insights into underlying molecular mechanisms and potential targets for therapeutic intervention.

3.3.4. Drug Discovery and Repurposing

AI-driven virtual screening methods, such as molecular docking and AI-based drug design algorithms, expedite the identification of candidate compounds targeting dysregulated pathways in CDG. By harnessing computational approaches, AI accelerates drug discovery efforts and facilitates the repurposing of existing medications for CDG treatment, addressing the urgent need for effective therapeutic options.

3.3.5. Clinical Decision Support Systems (CDSS)

AI-powered Clinical Decision Support Systems (CDSS) integrate Natural Language Processing (NLP) techniques to extract and analyze clinical data from Electronic Health Records (EHRs) of CDG patients.[31] Machine Learning algorithms applied to this data enable early disease detection, personalized treatment planning, and real-time monitoring of disease progression, thereby improving clinical outcomes and patient management strategies.[32]

3.3.6. Imaging Analysis for Phenotyping

Utilizing Computer Vision and image analysis algorithms, AI techniques analyze medical imaging data (e.g., MRI, CT scans) to detect characteristic phenotypic features associated with CDG-related organ abnormalities.[33] By automating phenotypic characterization, AI enhances diagnostic precision and aids in monitoring disease progression, facilitating timely interventions and personalized patient care.

3.3.7. Biochemical Signature Identification

AI methodologies applied to metabolomics data and spectroscopic techniques, such as Infrared (IR) spectroscopy, identify unique biochemical signatures indicative of CDG subtypes.[34] These AI-driven approaches elucidate metabolic dysregulations underlying CDG pathophysiology, offering diagnostic biomarkers and informing therapeutic strategies aimed at restoring glycosylation homeostasis.

4. AI for Congenital Disorders of Glycosylation (CDG)

4.1. AI for CDG Disease Mechanisms Elucidation

- Prediction of Glycosylation Sites The prediction of glycosylation sites within proteins is crucial for understanding CDG disease mechanisms.[35] A glycosylation prediction program (GPP) employing an RF algorithm and pairwise patterns has been developed to automatically identify putative glycosylation sites in glycoprotein sequences.[36] Additionally, ensembles of SVM classifiers have been utilized for this purpose, aiming to enhance the accuracy of glycosylation site prediction. These advancements in glycosylation site prediction hold significant promise for guiding experimental investigations and improving CDG patients' quality of life.[37]
- Identification of Golgi Proteins the Golgi apparatus plays a pivotal role in glycosylation, and defects in Golgi proteins can lead to CDG. [37] To advance diagnosis and basic research in CDG, the identification of sub-Golgi protein types (isGPT) model has been developed using RF and SVM. This model accurately identifies trans- and cis-Golgi proteins, contributing to a better understanding of CDG pathogenesis and facilitating targeted drug development.[38]

4.2. AI for CDG Diagnosis, Classification, and Characterization

Accurate and timely diagnosis is challenging in CDG. AI tools such as Face2Gene and the phenotype-based rare disease auxiliary diagnosis (RDAD) system have been employed to assist clinicians in diagnosing PMM2-CDG, the most common N-glycosylation disease. These tools successfully identified recognizable facial patterns in PMM2-CDG patients and ranked the most likely candidate RDs, aiding in disease diagnosis and natural history definition. [39,40] Additionally, a pioneering ML model has been utilized to classify EXT1-CDG and EXT2-CDG, providing insights into disease pathogenesis and facilitating diagnosis. Furthermore, predictive models based on electronic health records (EHRs) have been developed to predict young stroke, a common manifestation of CDG, contributing to improved disease management.

4.3. AI for Therapy Discovery in CDG

Therapy discovery in CDG is challenging, but AI-driven approaches hold promise for identifying potential treatment options. mRNA modeling strategies have been explored to repurpose drugs for CDG disorders, utilizing tools like the Connectivity Map and causal reasoning engine to analyze mRNA responsiveness to drugs in various RDs, including SLC38A9-CDG, MAN2B1-CDG, and PMM2-CDG. Excitingly, these analyses have identified putative drug-gene interactions, offering potential avenues for therapeutic intervention in CDG. [41,42]

In conclusion, AI-driven approaches are revolutionizing our understanding of CDG disease mechanisms, aiding in diagnosis and classification, and offering promising avenues for therapeutic discovery, ultimately improving the management and quality of life for CDG patients.

4.4. Comparison with Existing Literature

Our review of the literature on Artificial Intelligence (AI) applications in rare diseases (RDs) aligns with existing research while also providing novel insights and perspectives. Previous studies have highlighted the potential of AI-driven approaches in various aspects of RD research, including diagnosis, elucidation of disease mechanisms, therapeutic approaches, and management of patient health records.[43,44] Similarly, our findings underscore the transformative impact of AI technologies across these domains, emphasizing their ability to expedite and enhance the

accuracy of RD diagnosis through Medical Decision Support Systems (DSS) and deep learning models like DeepGestalt. Moreover, our review corroborates existing literature by showcasing the utility of AI methodologies, such as machine learning algorithms and multi-omics integrative approaches, in deciphering complex disease mechanisms and identifying potential therapeutic targets.[45]

In comparing with existing literature, our review contributes novel insights by synthesizing recent advancements in AI applications for RDs. While previous studies have demonstrated the efficacy of AI-driven approaches in drug repositioning, clinical trial optimization, and biomarker discovery, our review provides a comprehensive overview of these methodologies and their implications for RD research and patient care.[46] Furthermore, our analysis of AI-driven approaches in patient health registries and medical records management builds upon existing literature by highlighting the role of Natural Language Processing (NLP) and Named Entity Recognition (NER) in extracting and analyzing unstructured clinical data. By contextualizing our findings within the broader landscape of AI applications in RD research, our review offers a nuanced understanding of the current state of the field and identifies areas for future research and interdisciplinary collaboration.[47]

In summary, while our review aligns with existing literature in recognizing the transformative potential of AI technologies in RD research, it also extends this understanding by providing a comprehensive synthesis of recent advancements and highlighting areas for further exploration. By comparing and contextualizing our findings with existing research, we contribute to the ongoing dialogue on the role of AI in improving outcomes for RD patients and underscore the importance of continued research and collaboration in harnessing the full potential of AI-driven approaches in rare disease research and clinical practice.

4.5. Case Studies of AI Success in Treating Rare Diseases

4.5.1. AI Applications in Rare Diseases

AI, particularly deep learning, is successfully applied in basic research, diagnosis, drug discovery, and clinical trials for rare diseases.[48]

Only 5% of over 7000 rare diseases worldwide have a treatment, highlighting the potential for AI to fill this gap.[49]

AI technologies can integrate and analyze data from various sources to overcome challenges like low diagnostic rates and geographical dispersion in rare diseases [50]

4.5.2. Specific Applications of AI in Rare Diseases

AI has been used to identify disease biomarkers, increase patient recruitment for clinical trials, and discover drugs for repurposing in rare diseases.

AI technologies can assess patient experiences, improve recruitment and engagement through social media, and monitor patient adherence in clinical trials.[51]

AI algorithms for mutation detection, prediction, and classification can enhance diagnosis rates and uncover new disease mechanisms and therapeutic targets in rare diseases.

4.5.3. AI in Congenital Disorders of Glycosylation (CDG)

CDG, a group of orphan rare diseases, serves as a potential study model for other common diseases and rare diseases.

Exciting developments in AI have shown great potential for revolutionizing the therapeutic landscape of CDG.[52]

4.5.4. AI and Therapeutic Development

The use of AI algorithms in diagnosis and drug discovery has significantly benefited biomedicine, including rare diseases.

AI tools applicable to a wide range of diseases, including rare diseases, have been developed for diagnosis, drug discovery, and preclinical research.

AI's role in biomedicine continues to evolve, with a focus on patient-centered approaches, including diagnosis, disease classification, therapeutic approaches, and patient registries.[53]

In conclusion, AI has shown promising results in treating rare diseases by enhancing diagnosis, drug discovery, patient engagement, and therapeutic development. The use of AI technologies in rare diseases, such as CDG, demonstrates the potential for significant advancements in overcoming the challenges associated with these conditions.

4.6. AI Applications in CDG Research and Diagnosis

The application of artificial intelligence (AI) tools in Congenital Disorders of Glycosylation (CDG) research and diagnosis has shown significant promise. AI algorithms, such as PredictSNP and Face2Gene, have demonstrated enhanced accuracy in predicting genetic variant pathogenicity and identifying characteristic phenotypic features associated with CDG, respectively.[56] These advancements have facilitated early and precise diagnosis, thereby improving patient outcomes. Additionally, AI-driven approaches in predicting glycosylation sites and identifying key Golgi proteins have provided valuable insights into CDG disease mechanisms, furthering our understanding of this complex disorder.[58]

4.6.1. Challenges and Opportunities

Despite the remarkable progress in AI-driven CDG research and diagnosis, several challenges persist. Data quality, model interpretability, and ethical considerations are among the key challenges that need to be addressed to fully realize the potential of AI in CDG management. Integrating diverse data sources, enhancing model transparency, and implementing robust ethical frameworks are essential steps towards overcoming these challenges.[57] Furthermore, collaborative efforts between researchers, clinicians, and patient advocacy groups are crucial for advancing AI-driven solutions in CDG research and treatment.[59,60]

4.6.2. Future Directions

Looking ahead, the continued evolution of AI technologies holds great promise for CDG and other rare genetic disorders.[61] Integrative approaches that combine multi-omics data, advanced machine learning algorithms, and ethical data practices will further enhance our understanding and management of CDG. Moreover, expanding access to AI-driven diagnostic tools and therapeutic discovery platforms will democratize CDG care, ensuring equitable access to innovative solutions. By addressing existing challenges and embracing collaborative approaches, we can harness the full potential of AI to revolutionize CDG management and improve patient outcomes [62,63]

5. Conclusion

AI holds immense potential in transforming the landscape of CDG research and treatment. By enhancing diagnostics, personalizing treatment, and accelerating drug discovery, AI can bring new hope to patients with CDG and other rare diseases. Collaborative efforts among researchers, clinicians, and policymakers are essential to harness the full power of AI and achieve breakthroughs in rare disease frontiers.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Freeze, H. H. (2013). Understanding human glycosylation disorders: biochemistry leads the charge. The Journal of Biological Chemistry, 288(10), 6936-6945.

[2] Sernadela, P.; González-Castro, L.; Carta, C.; van der Horst, E.; Lopes, P.; Kaliyaperumal, R.; Thompson, M.; Thompson, R.; Queralt-Rosinach, N.; Lopez, E.; et al. Linked Registries: Connecting Rare Diseases Patient Registries through a Semantic Web Layer. BioMed Res. Int. **2017**, 2017, 1–13. [CrossRef]

[3] Ekins, S. Industrializing rare disease therapy discovery and development. Nat. Biotechnol. **2017**, 35, 117–118. [CrossRef] [PubMed]

[4] About Rare Diseases. Available online: https://www.eurordis.org/about-rare-diseases (accessed on 18 September 2019).

[5] Ronicke, S.; Hirsch, M.C.; Turk, E.; Larionov, K.; Tientcheu, D.; Wagner, A.D. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. Orphanet J. Rare Dis. **2019**, 14, 69. [CrossRef] [PubMed]

[6] Nestler-Parr, S.; Korchagina, D.; Toumi, M.; Pashos, C.L.; Blanchette, C.; Molsen, E.; Morel, T.; Simoens, S.; Kaló, Z.; Gatermann, R.; et al. Challenges in Research and Health Technology Assessment of Rare Disease Technologies: Report of the ISPOR Rare Disease Special Interest Group. Value Health **2018**, 21, 493–500. [CrossRef] [PubMed]

[7] Stoller, J.K. The Challenge of Rare Diseases. Chest **2018**, 153, 1309–1314. [CrossRef]

[8] Ahmed, M.A.; Okour, M.; Brundage, R.; Kartha, R.V. Orphan drug development: The increasing role of clinical pharmacology. J. Pharmacokinet. Pharmacodyn. **2019**, 46, 395–409. [CrossRef]

[9] Kaufmann, P.; Pariser, A.R.; Austin, C. From scientific discovery to treatments for rare diseases—The view from the National Center for Advancing Translational Sciences—O_ce of Rare Diseases Research. Orphanet J. Rare Dis. **2018**, 13, 196. [CrossRef]

[10] Schee genannt Halfmann, S.; Mählmann, L.; Leyens, L.; Reumann, M.; Brand, A. Personalized Medicine: What's in it for Rare Diseases? In Rare Diseases Epidemiology: Update and Overview; Posada de la Paz, M., Taruscio, D., Groft, S.C., Eds.; Springer International Publishing: Cham, Switzerland, 2017; Volume 1031, pp. 387–404, ISBN 978-3-319-67142-0.

[11] Toh, T.S.; Dondelinger, F.; Wang, D. Looking beyond the hype: Applied AI and machine learning in translational medicine. EBioMedicine **2019**, 47, 607–615. [CrossRef]

[12] Bishop, C.M. Pattern Recognition and Machine Learning; Information Science and Statistics; Springer: New York, NY, USA, 2006; ISBN 978-0-387-31073-2.

[13] Pasa, F.; Golkov, V.; Pfei_er, F.; Cremers, D.; Pfei_er, D. E_cient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization. Sci. Rep. **2019**, 9, 6268. [CrossRef]

[14] Portnoi, T.; Yala, A.; Schuster, T.; Barzilay, R.; Dontchos, B.; Lamb, L.; Lehman, C. Deep Learning Model to Assess Cancer Risk on the Basis of a Breast MR Image Alone. Am. J. Roentgenol. **2019**, 213, 227–233. [CrossRef]

[15] Alashwal, H.; El Halaby, M.; Crouse, J.J.; Abdalla, A.; Moustafa, A.A. The Application of Unsupervised Clustering Methods to Alzheimer's Disease. Front. Comput. Neurosci. **2019**, 13, 31. [CrossRef] [PubMed]

[16] Yang, M.-H.; Yang, F.-Y.; Oyang, Y.-J. Application of density estimation algorithms in analyzing co-morbidities of migraine. Netw. Model. Anal. Health Inform. Bioinform. **2013**, 2, 95–107. [CrossRef] [PubMed]

[17] Lunceford, J.K.; Davidian, M.; Tsiatis, A.A. Estimation of Survival Distributions of Treatment Policies in Two-Stage Randomization Designs in Clinical Trials. Biometrics **2002**, 58, 48–57. [CrossRef] [PubMed]

[18] Panuccio, G.; Guez, A.; Vincent, R.; Avoli, M.; Pineau, J. Adaptive control of epileptiform excitability in an in vitro model of limbic seizures. Exp. Neurol. **2013**, 241, 179–183. [CrossRef] [PubMed]

[19] Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. Nat. Med. **2019**, 25, 44–56. [CrossRef]

[20] Sayers, E. A General Introduction to the E-utilities; National Center for Biotechnology Information (US): Bethesda, MD, USA, 2010.

[21] Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kau_, F.; Wilczynski, B.; et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics **2009**, 25, 1422–1423. [CrossRef] Genes **2019**, 10, 978 20 of 24

[22] Carter, H.; Douville, C.; Stenson, P.D.; Cooper, D.N.; Karchin, R. Identifying Mendelian disease genes with the Variant E_ect Scoring Tool. BMC Genom. **2013**, 14, S3. [CrossRef]

[23] Alirezaie, N.; Kernohan, K.D.; Hartley, T.; Majewski, J.; Hocking, T.D. ClinPred: Prediction Tool to Identify Disease-Relevant Nonsynonymous Single-Nucleotide Variants. Am. J. Hum. Genet. **2018**, 103, 474–483. [CrossRef]

[24] Bosio, M.; Drechsel, O.; Rahman, R.; Muyas, F.; Rabionet, R.; Bezdan, D.; Domenech Salgado, L.; Hor, H.; Schott, J.; Munell, F.; et al. eDiVA—Classification and prioritization of pathogenic variants for clinical diagnostics. Hum. Mutat. **2019**, 40, 865–878. [CrossRef]

[25] Sundaram, L.; Gao, H.; Padigepati, S.R.; McRae, J.F.; Li, Y.; Kosmicki, J.A.; Fritzilas, N.; Hakenberg, J.; Dutta, A.; Shon, J.; et al. Predicting the clinical impact of human mutation with deep neural networks. Nat. Genet. **2018**, 50, 1161–1170. [CrossRef]

[26] Dehiya, V.; Thomas, J.; Sael, L. Impact of structural prior knowledge in SNV prediction: Towards causal variant finding in rare disease. PLoS ONE **2018**, 13, e0204101. [CrossRef] [PubMed]

[27] Buske, O.J.; Manickaraj, A.; Mital, S.; Ray, P.N.; Brudno, M. Identification of deleterious synonymous variants in human genomes. Bioinformatics **2013**, 29, 1843–1850. [CrossRef] [PubMed]

[28] Browne, C.; Timson, D.J. In Silico Prediction of the E_ects of Mutations in the Human Mevalonate Kinase Gene: Towards a Predictive Framework for Mevalonate Kinase Deficiency: E_ects of Mutations in Human MVK. Ann. Hum. Genet. **2015**, 79, 451–459. [CrossRef] [PubMed]

[29] Orange, J.S.; Glessner, J.T.; Resnick, E.; Sullivan, K.E.; Lucas, M.; Ferry, B.; Kim, C.E.; Hou, C.; Wang, F.; Chiavacci, R.; et al. Genome-wide association identifies diverse causes of common variable immunodeficiency. J.Allergy Clin. Immunol. **2011**, 127, 1360–1367.e6. [CrossRef] [PubMed]

[30] Jaganathan, K.; Kyriazopoulou Panagiotopoulou, S.; McRae, J.F.; Darbandi, S.F.; Knowles, D.; Li, Y.I.; Kosmicki, J.A.; Arbelaez, J.; Cui, W.; Schwartz, G.B.; et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell **2019**, 176, 535–548.e24. [CrossRef]

[31] Papadimitriou, S.; Gazzo, A.; Versbraegen, N.; Nachtegael, C.; Aerts, J.; Moreau, Y.; Van Dooren, S.; Nowé, A.; Smits, G.; Lenaerts, T. Predicting disease-causing variant combinations. Proc. Natl. Acad. Sci. USA **2019**, 116, 11878–11887. [CrossRef]

[32] Yang, X.; Song, Z.; Wu, C.; Wang, W.; Li, G.; Zhang, W.; Wu, L.; Lu, K. Constructing a database for the relations between CNV and human genetic diseases via systematic text mining. BMC Bioinform. **2018**, 19, 528. [CrossRef]

[33] Boudellioua, I.; Kulmanov, M.; Schofield, P.N.; Gkoutos, G.V.; Hoehndorf, R. DeepPVP: Phenotype-based prioritization of causative variants using deep learning. BMC Bioinform. **2018**, 20, 65. [CrossRef]

[34] Li, Q.; Zhao, K.; Bustamante, C.D.; Ma, X.; Wong, W.H. Xrare: A machine learning method jointly modelling phenotypes and genetic evidence for rare disease diagnosis. Genet. Med. **2019**, 21, 2126–2134. [CrossRef]

[35] Rao, A.; Vg, S.; Joseph, T.; Kotte, S.; Sivadasan, N.; Srinivasan, R. Phenotype-driven gene prioritization for rare diseases using graph convolution on heterogeneous networks. BMC Med. Genom. **2018**, 11, 57. [CrossRef]

[36] Paul, R.; Groza, T.; Hunter, J.; Zankl, A. Inferring characteristic phenotypes via class association rule mining in the bone dysplasia domain. J. Biomed. Inform. **2014**, 48, 73–83. [CrossRef] [PubMed]

[37] Ehsani-Moghaddam, B.; Queenan, J.A.; MacKenzie, J.; Birtwhistle, R.V. Mucopolysaccharidosis type II detection by Naïve Bayes Classifier: An example of patient classification for a rare disease using electronic medical records from the Canadian Primary Care Sentinel Surveillance Network. PLoSONE**2018**, 13, e0209018. [CrossRef] [PubMed]

[38] Shen, F.; Liu, S.; Wang, Y.; Wen, A.; Wang, L.; Liu, H. Utilization of Electronic Medical Records and Biomedical Literature to Support the Diagnosis of Rare Diseases Using Data Fusion and Collaborative Filtering Approaches. JMIR Med. Inform. **2018**, 6, e11301. [CrossRef]

[39] Shen, F.; Zhao, Y.;Wang, L.; Mojarad, M.R.;Wang, Y.; Liu, S.; Liu, H. Rare disease knowledge enrichment through a data-driven approach. BMC Med Inform. Decis. Mak. **2019**, 19, 32. [CrossRef]

[40] Javed, A.; Agrawal, S.; Ng, P.C. Phen-Gen: Combining phenotype and genotype to analyze rare disorders. Nat. Methods **2014**, 11, 935–937. [CrossRef]

[41] Garcelon, N.; Neuraz, A.; Salomon, R.; Faour, H.; Benoit, V.; Delapalme, A.; Munnich, A.; Burgun, A.; Rance, B. A clinician friendly data warehouse oriented toward narrative reports: Dr. Warehouse. J. Biomed. Inform. **2018**, 80, 52–63. [CrossRef] Genes **2019**, 10, 978 21 of 24

[42] Garcelon, N.; Neuraz, A.; Benoit, V.; Salomon, R.; Kracker, S.; Suarez, F.; Bahi-Buisson, N.; Hadj-Rabia, S.; Fischer, A.; Munnich, A.; et al. Finding patients using similarity measures in a rare diseases-oriented clinical data warehouse: Dr. Warehouse and the needle in the needle stack. J. Biomed. Inform. **2017**, 73, 51–61. [CrossRef]

[43] Smpokou, P.; Lanpher, B.C.; Rosenbaum, K.N. Important Considerations in the Initial Clinical Evaluation of the Dysmorphic Neonate. Adv. Neonatal Care **2015**, 15, 248–252. [CrossRef]

[44] Smith, K.; Piccinini, F.; Balassa, T.; Koos, K.; Danka, T.; Azizpour, H.; Horvath, P. Phenotypic Image Analysis Software Tools for Exploring and Understanding Big Image Data from Cell-Based Assays. Cell Syst. **2018**, 6, 636–653. [CrossRef]

[45] Gurovich, Y.; Hanani, Y.; Bar, O.; Nadav, G.; Fleischer, N.; Gelbman, D.; Basel-Salmon, L.; Krawitz, P.M.; Kamphausen, S.B.; Zenker, M.; et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat. Med. **2019**, 25, 60–64. [CrossRef]

[46] LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature **2015**, 521, 436–444. [CrossRef] [PubMed]

[47] Zarate, Y.A.; Smith-Hicks, C.L.; Greene, C.; Abbott, M.-A.; Siu, V.M.; Calhoun, A.R.U.L.; Pandya, A.; Li, C.; Sellars, E.A.; Kaylor, J.; et al. Natural history and genotype-phenotype correlations in 72 individuals with SATB2 -associated syndrome. Am. J. Med. Genet. **2018**, 176, 925–935. [CrossRef] [PubMed]

[48] Basel-Vanagaite, L.; Wolf, L.; Orin, M.; Larizza, L.; Gervasini, C.; Krantz, I.D.; Deardo_, M.A. Recognition of the Cornelia de Lange syndrome phenotype with facial dysmorphology novel analysis: Recognition of the CdLS Phenotype with FDNA. Clin. Genet. **2016**, 89, 557–563. [CrossRef] [PubMed]

[49] Liehr, T.; Acquarola, N.; Pyle, K.; St-Pierre, S.; Rinholm, M.; Bar, O.; Wilhelm, K.; Schreyer, I. Next generation phenotyping in Emanuel and Pallister-Killian syndrome using computer-aided facial dysmorphology analysis of 2D photos: LIEHR et al. Clin. Genet. **2018**, 93, 378–381. [CrossRef] [PubMed]

[50] Marbach, F.; Rustad, C.F.; Riess, A.; _uki´c, D.; Hsieh, T.-C.; Jobani, I.; Prescott, T.; Bevot, A.; Erger, F.; Houge, G.; et al. The Discovery of a LEMD2-Associated Nuclear Envelopathy with Early Progeroid Appearance Suggests Advanced Applications for AI-Driven Facial Phenotyping. Am. J. Hum. Genet. **2019**, 104, 749–757. [CrossRef]

[51] Learned-Miller, E.; Lu, Q.; Paisley, A.; Trainer, P.; Blanz, V.; Dedden, K.; Miller, R. Detecting Acromegaly:Screening for Disease with a Morphable Model. In Medical Image Computing and Computer-Assisted Intervention—MICCAI 2006; Larsen, R., Nielsen, M., Sporring, J., Eds.; Springer: Heidelberg, Germany, 2006; Volume 4191, pp. 495–503. ISBN 978-3-540-44727-6.

[52] Spinelli, E.G.; Mandelli, M.L.; Miller, Z.A.; Santos-Santos, M.A.; Wilson, S.M.; Agosta, F.; Grinberg, L.T.; Huang, E.J.; Trojanowski, J.Q.; Meyer, M.; et al. Typical and atypical pathology in primary progressive aphasia variants: Pathology in PPA Variants. Ann Neurol. **2017**, 81, 430–443. [CrossRef]

[53] Hsieh, T.-C.; Mensah, M.A.; Pantel, J.T.; Aguilar, D.; Bar, O.; Bayat, A.; Becerra-Solano, L.; Bentzen, H.B.; Biskup, S.; Borisov, O.; et al. PEDIA: Prioritization of exome data by image analysis. Genet. Med. **2019**. [CrossRef]

[54] Lux, A.; Müller, R.; Tulk, M.; Olivieri, C.; Zarrabeita, R.; Salonikios, T.; Wirnitzer, B. HHT diagnosis by Mid-infrared spectroscopy and artificial neural network analysis. Orphanet J. Rare Dis. **2013**, 8, 94. [CrossRef] [PubMed]

[55] Han, I.; Kim, J.H.; Park, H.; Kim, H.-S.; Seo, S.W. Deep learning approach for survival prediction for patients with synovial sarcoma. Tumour Biol. **2018**, 40, 101042831879926. [CrossRef]

[56] Van IJzendoorn, D.G.P.; Szuhai, K.; Briaire-de Bruijn, I.H.; Kostine, M.; Kuijjer, M.L.; Bovée, J.V.M.G. Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. PLoS Comput. Biol. **2019**, 15, e1006826. [CrossRef]

[57] Noto, K.; Majidi, S.; Edlow, A.G.; Wick, H.C.; Bianchi, D.W.; Slonim, D.K. CSAX: Characterizing Systematic Anomalies in eXpression Data. J. Comput. Biol. **2015**, 22, 402–413. [CrossRef] [PubMed]

[58] Taroni, J.N.; Grayson, P.C.; Hu, Q.; Eddy, S.; Kretzler, M.; Merkel, P.A.; Greene, C.S. MultiPLIER: A Transfer Learning Framework for Transcriptomics Reveals Systemic Features of Rare Disease. Cell Syst. **2019**, 8, 380–394.e4. [CrossRef] [PubMed]

[59] DeAndrés-Galiana, E.J.; Fernández-Martínez, J.L.; Sonis, S.T. Design of Biomedical Robots for Phenotype Prediction Problems. J. Comput. Biol. **2016**, 23, 678–692. [CrossRef] [PubMed]

[60] DeAndrés-Galiana, E.J.; Fernández-Martínez, J.L.; Sonis, S.T. Sensitivity analysis of gene ranking methods in phenotype prediction. J. Biomed. Inform. **2016**, 64, 255–264. [CrossRef] [PubMed] Genes **2019**, 10, 978 22 of 24

[61] Romero, R.; Ramanathan, A.; Yuen, T.; Bhowmik, D.; Mathew, M.; Munshi, L.B.; Javaid, S.; Bloch, M.; Lizneva, D.; Rahimova, A.; et al. Mechanism of glucocerebrosidase activation and dysfunction in Gaucher disease unraveled by molecular dynamics and deep learning. Proc. Natl. Acad. Sci. USA **2019**, 116, 5086–5095. [CrossRef]

[62] Danter,W.R. DeepNEU: Cellular reprogramming comes of age – a machine learning platform with application to rare diseases research. Orphanet J. Rare Dis. **2019**, 14, 13. [CrossRef]

[63] Hoehndorf, R.; Schofield, P.N.; Gkoutos, G.V. Analysis of the human diseasome using phenotype similarity between common, genetic and infectious diseases. Sci. Rep. **2015**, 5, 10888. [CrossRef]

[64] Lee, Y.; Krishnan, A.; Oughtred, R.; Rust, J.; Chang, C.S.; Ryu, J.; Kristensen, V.N.; Dolinski, K.; Theesfeld, C.L.; Troyanskaya, O.G. A Computational Framework for Genome-wide Characterization of the Human Disease Landscape. Cell Syst. **2019**, 8, 152–162. [CrossRef]

[65] Crow, R.A.; Hart, K.A.; McDermott, M.P.; Tawil, R.; Martens, W.B.; Herr, B.E.; McColl, E.; Wilkinson, J.; Kirschner, J.; King, W.M.; et al. A checklist for clinical trials in rare disease: Obstacles and anticipatory actions—lessons learned from the FOR-DMD trial. Trials **2018**, 19, 291. [CrossRef]