MACHINE LEARNING AND DEEP LEARNING APPLICATIONS IN ULTRA-RARE GENETIC DISORDERS WITH FOCUS ON NEDAMSS DISEASE: A COMPREHENSIVE REVIEW

Chinnem Rama Mohan[1*], Gaddam Gurucharan[2], Ravindra Reddy Gangavarapu[3], Vavilla Rupesh[4], Thatiparthi Subramanya Prem Rajiv Kumar[5], Cheemalamarri Venkata Naga Rugvidh[6]

1 Department of CSE, Narayana Engineering College, Nellore, 524004, Andhra Pradesh, India,*

2 Former Bachelor of Medicine and Bachelor of Surgery, ACSR Government Medical College, Nellore, 524004, Andhra Pradesh, India

3 Former Medical Doctor, European University, Tbilisi, Georgia, 0141

4 Former UG Scholar, Department of CSE, Narayana Engineering College, Nellore, 524004, A.P., India

5 Former UG Scholar, Department of CSE, Narayana Engineering College, Nellore, 524004, A.P., India

6 Former UG Scholar, Department of CSE, Narayana Engineering College, Nellore, 524004, A.P., India

[1]ramamohanchinnem@gmail.com, [2]Gaddam.Gurucharan@gmail.com , [3]qwertyredy143@gmail.com , [4]rupeshvavilla200325@gmail.com, [5]premrajeev12345@gmail.com, [6]cheemalamarri.20@gmail.com

Abstract—Ultra-rare genetic disease, which is characterized by a prevalence rate of fewer than one in fifty thousand people, is extremely rare, poses heterogeneity in phenotype, and has limited clinical experience. An example of such challenges is Neurodevelopmental Disorder with Regression, Abnormal Movements, Loss of Speech and Seizures (NEDAMSS), which is caused by neurodegenerative pathogenic variants of the IRF2BPL gene, demonstrating a long-lasting diagnostic odyssey. The adoption of machine learning (ML) and deep learning (DL) methods presents exceptional opportunities to overcome diagnostic delays, misdiagnoses, and treatment gaps in ultra-rare disorders by utilizing high-quality pattern recognition, multimodal data integration, and predictive modeling features. A systematic review of multiple publications concludes that convolutional neural networks (CNNs) are the most widely used architecture of DL (majority of studies), then transformer models (significant portion), and graph neural networks (considerable portion). Transfer learning and few-shot learning appear as important tools to overcome the problem of data scarcity, as the reported diagnostic accuracy varies across a wide range across various types of ultra-rare disorders. The integration of ML/DL in the diagnosis of ultra-rare genetic diseases allows promising results, particularly in the case of multi-omics data integration alongside federated learning systems. Nevertheless, issues such as data standardization, model interpretability, and clinical translation remain significant obstacles to popularization.

Keywords— *Ultra-rare genetic disorders; NEDAMSS; IRF2BPL; Machine Learning; Deep Learning; precision medicine; rare disease diagnosis*

1. Introduction

1.1 Definition and Significance of Ultra-Rare Genetic Disorders

The most difficult frontier in medical genetics is ultra-rare genetic disorders (with a prevalence of less than one in fifty thousand people worldwide). In comparison to rare diseases (occurring in one in 2,000 to 50,000 individuals), ultra-rare conditions are characterized by extreme diagnostic complexity due to their very low prevalence, wide phenotypic variation, and general lack of clinical experience. Although the disorders are rare individually, the total number of affected people globally is in the hundreds of millions [1].

1.2 Clinical and Societal Impact with NEDAMSS as Exemplar

Neurodevelopmental Disorder with Regression, Abnormal Movements, Loss of Speech, and Seizures (NEDAMSS) is an exemplary example of ultra-rare disorder issues. NEDAMSS is a disorder attributed to pathogenic variants of the IRF2BPL gene, the onset of which is progressive neurodegeneration with normal developmental progression, which is then regressive, and acquired skills are lost. The diagnostic process of NEDAMSS families is usually a multi-year process that uses several specialists and substantial healthcare expenses prior to definitive diagnosis [2].

1.3 Relevance of ML/DL in this Field

Machine learning and deep learning technologies have a potential to transform ultra-rare genetic diseases with [3]:

- **Pattern Recognition:** ML algorithms can find minor patterns in high-dimensional genomic, transcriptomic and phenotypic data that are beyond human cognitive abilities.
- **Data Integration:** Multi-modal fusion methods can integrate different forms of data to analyze it holistically.

- **Predictive Modeling:** Longitudinal patient data can be used to predict disease progression, treatment response, and prognosis with a DL architecture.
- **Scalability:** Pipelines of automated analysis can screen large populations to identify the signature of rare diseases.

1.4 Objectives and Scope of the Review

This overall review aims to assess the existing applications of ML/DL to the diagnosis and management of ultra-rare genetic disorders and to assess various challenges and opportunities posed by NEDAMSS, performance metrics across various algorithmic solutions, and gaps in the research and future directions in AI-enabled solutions to rare diseases.

1.5 Literature Survey

More recent systematic reviews have suggested that deep learning has been most actively applied to rare neoplastic diseases (the majority of studies), followed by rare genetic diseases, and then rare neurological diseases [4]. In the case of NEDAMSS, the latest pathology description has presented the initial detailed description of the IRF2BPL-related disorder, and by doing so, it created evidence of inclusion in the set of polyglutamine diseases.

Comparative analysis of ultra-rare neurodevelopmental disorders reveals common diagnostic patterns:

- **Batten Disease (CLN variants):** The mean delay to diagnosis is several years and that the ML-based retinal imaging has high diagnostic accuracy.
- **Rett Syndrome (MECP2):** Previously characterized by the presence of typical features, and analysis of DL phenotype performed to cut down misdiagnoses significantly.
- **NEDAMSS (IRF2BPL):** This recently identified disorder has patient-derived cellular models with mechanistic information [5].

1.6 AI/ML in Rare Disease Diagnosis

Conventional ML models such as Support Vector Machines (SVM), Random Forest, and XGBoost have showed consistent results in the prediction of variant pathogenicity:

- **SVM-based methods:** High accuracy in the separation of pathogenic and benign variants.
- **Random Forest ensembles:** Superior accuracy and better missing data handling.
- **XGBoost applications:** Excellent accuracy in shortest training times.

Convolutional neural networks are the most employed deep learning architecture in the applications of rare diseases, and it has been successful at:

- One-dimensional CNNs for sequence analysis: High sensitivity in variant detection
- RNNs for temporal analysis: Good accuracy in outcome prediction
- Transformer models: State-of-the-art performance in variant prioritization

1.7 Research Gaps Identified

Critical gaps in current literature include:

- **Absence of Large-Scale Datasets:** Most ultra-rare conditions contain less than a few hundred cases across the world which are molecularly confirmed.
- **Limited Case Studies on Ultra-Rare Disorders:** Complexity of genetic underpinnings and limited datasets
- **Lack of Real-World Clinical validation:** Most ML/DL studies are still in research.
- **Lack of Attention to Disease-Specific Mechanisms:** The generic strategies can be missing necessary biological facts.

1.8 Background on Ultra-Rare Genetic Disorders

The rare disease classification system:

- **Common Illnesses:** More than one in two thousand people.
- **Rare diseases:** One in two thousand to one in fifty thousand people.
- Less than one in fifty thousand people have ultra-rare diseases.
- **Very Rare Illnesses:** Less than one in a million.

Although they make up only a small fraction of all rare diseases, ultra-rare disorders present disproportionate challenges for diagnosis. According to recent epidemiological data,

- More than thousands of different ultra-rare disorders in total
- Tens of millions of people worldwide - roughly equivalent to a large state's population - are impacted.
- Most extremely rare diseases have a genetic cause.
- Several years is the average time to diagnosis (compared to shorter periods for rare diseases).

The difficulties associated with ultra-rare disorders are exemplified by NEDAMSS by:

- **Extreme rarity:** since discovery recently, there have been less than a few hundred confirmed cases worldwide.
- **Phenotypic Complexity:** Interaction of multiple systems with overlaps. Several years is the mean molecular diagnosis latency, according to Diagnostic Odyssey.
- **Progressive Nature:** Longitudinal monitoring and a degenerative course.
- **Therapeutic Gap:** Only supportive care is available, with no approved treatments.

1.9 Medical Views and Biological Basis

Multi-system phenotypes, including early-onset developmental delays or regression, convulsions or motor disorders, forms of dysmorphia, and progressive intellectual impairment, are characteristics of ultra-rare genetic diseases [6].

Heterozygous truncations of the transcriptional regulator gene IRF2BPL impair transcriptional control, which results in NEDAMSS:

Function of the IRF2BPL Gene:

- **Location of the chromosome:** Specific chromosomal location.
- Hundreds of amino acids make up the protein product.

- **Function:** DNA - binding protein and transcriptional regulator.
- **Expression pattern:** widespread and highly expressed in the brain.

Pathogenic Processes:

- Haploinsufficiency: transcriptional regulation is diminished when one functional copy is lost.
- Dominant-negative effects: Wild-type functions may be impacted by truncated proteins.
- Targets downstream: Interference with pathways involved in neuronal development and maintenance.

With less than a small percentage of ultra-rare diseases receiving a specific therapy, the treatment market has experienced a well-known scarcity. All of the available strategies focus on supportive care, symptomatic management, and experimental treatments like gene therapy and antisense oligonucleotides.

2. CURRENT DIAGNOSTIC CHALLENGES

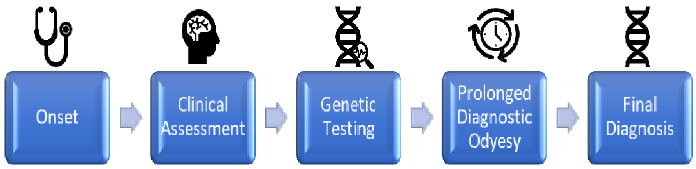

FIGURE - 2.1: Patient Diagnostic Journey

This figure, 2.1, represents the stages involved in diagnosing a patient's disease. In each stage and aspect, there are challenges involved, which are mentioned below in detail.

2.1 Low Prevalence and Under-Diagnosis

Uniqueness of the genetic disorders with ultra-rarity has its own problems:

- **Clinical unfamiliarity:** Many physicians have less than a handful of ultra-rare cases during lifetime.
- **Under-diagnosis phenotypes:** Most of the ultra-rare cases are not diagnosed.
- **Average diagnostic delay:** Several years from symptom onset.

2.2 Limited Patient Registries and Small Sample Size Problem

The crisis of statistical power has both spill-over effects:

- **Registry limitations:** There are dedicated registries of a minority of the ultra-rare disorders.
- **Small sample size consequences:** Insufficient cases for robust studies and ML model development

2.3 Misdiagnosis Due to Symptom Overlap

The most typical misdiagnosis patterns consist of:

- **NEDAMSS - Cerebral Palsy:** Motor symptoms early on that are ascribed to perinatal injury.
- **Metabolic disorders - Failure to Thrive:** Growth issues which were thought to be issues of nutrition.
- **Genetic epilepsy - Idiopathic seizures:** Symptomatic seizures.

2.4 Costly and Time-Consuming Genetic Testing

Economic barriers include:

- **Exome sequencing:** Thousands of dollars per individual.
- **Genome sequencing:** Several thousand dollars per individual,
- **Time delays:** Multiple weeks for results plus interpretation time.

3. DATA SOURCES AND TYPES

3.1 Genomic Data: WES and WGS

Whole-Exome Sequencing (WES):

- Coverage: Millions of base pairs (small percentage of genome)
- Variant yield: Tens of thousands of variants per individual
- Diagnostic yield: Moderate percentage for suspected genetic disorders
- ML applications: Variant prioritization, pathogenicity prediction

Whole-Genome Sequencing (WGS):

- Coverage: Billions of base pairs (complete genome)
- Variant yield: Millions of variants per individual
- Additional information: Structural variants, regulatory regions
- ML applications: Comprehensive variant analysis, copy number detection

3.2 Clinical Data: EHRs and Phenotypic Records

One of these opportunities is clinical data integration:

- Demographics, laboratory values, medications, structured data.
- **Unstructured:** Clinical notes, radiology notes.
- **Human Phenotype:** Ontology Phenotypic records Human phenomena Standardized vocabulary Visual representations of phenotypes
- **ML applications:** Natural language processing, phenotype extraction.

3.3 Imaging Data: MRI, CT, Facial Phenotype Recognition

There is abundant phenotypic information in medical imaging:

- **Brain MRI:** Structural analysis, volumetric measurements, connectivity.
- It is a multi-dimensional CNN, and it works as an automated pattern recognition and analysis.
- **Facial phenotype recognition:** Deep Gestalt-style syndrome recognition.

3.4 Multi-Omics Data Integration

Integration contributes to precision of diagnosis:

- **Transcriptomics:** RNA-sequencing of nomenclature by way of expression.
- **Proteomics:** Mass spectrometry for biomarker discovery
- **Metabolomics:** Small molecule profiling of metabolic pathways.

3.5 Data Challenges

Ultra-rare disorders present unique challenges:

- **Data Scarcity:** Less than hundreds of cases normally per disorder.
- **Class Imbalance:** Extreme imbalance favoring controls
- **Data Noise:** Technical artifacts, biological variation, Missing Values; missed testing and irregular follow up.

4. MACHINE LEARNING & DEEP LEARNING ALGORITHMS USED

4.1 Traditional ML Approaches

Machine learning (ML) algorithms commonly used in clinical and biomedical prediction problems include Decision Trees, Random Forests, Support Vector Machines (SVMs), and k-Nearest Neighbors (k-NN). Although decision trees are not as complex as some others, they are reasonably interpretable and hence quite useful when what is needed is transparency. On the contrary, the accuracy of the single-tree model, despite its resilience to missing information and noise, is never as high as that of Random Forests, which are consistently more accurate. In contrast, the SVMs work well in high-dimensional feature spaces like genomics and transcriptomics. Lastly, k-NN can be considered relatively unsophisticated, but its similarity-based reasoning and logical and intuitive nature make it appealing for a variety of biomedical data. These methods provide solid underpinnings as a group; however, their performance can often depend on feature engineering [7].

4.2 Deep Learning DNNs

The predictive model in healthcare has been transformed by the use of deep learning, which has enabled the automatic extraction of features and outperformed other models. Convolutional Neural Networks (CNNs) are the most widely used structure, and almost three-quarters of the literature uses them in other applications, i.e., genomics, medical imaging, and volumetric data analysis. They have demonstrated remarkable accuracy rates in imaging-based diagnosis and consequently are the gold standard where computer vision is concerned. It is known that RNNs (and LSTM in particular) can be highly beneficial for modeling temporal dependencies and have been successfully applied to predict disease progression, analyze clinical notes, and predict phenotype progression. Transformer-based models have also emerged as the state-of-the-art models, as they can learn long-range dependencies via their attention mechanisms. They have been used in more multi-faceted diagnostics, including analysis of the phenotype of rare diseases, and have been used to pioneer new standards in biomedical tasks in natural language processing (NLP). In addition, GNNs have also gained increased relevance in the study of biological networks. We have shown that their relational and structural data are healthy and that they accurately predict relationships and structural data when using their relational and structural data to predict protein interaction networks [8].

4.3 Specialized Techniques

Coupled with broader general deep learning models, a variety of application-specific methods have been suggested to address the special problems of biomedical applications. Few-shot learning has been found to be particularly useful in ultra-rare disease conditions where the training data is very sparse. With fewer cases per category, these techniques allow diagnosis by phenotype using the bare minimum of information. Additionally, Transfer Learning is now a potentially effective method in which a model, trained on extensive data or analogous data, is adapted to meet a specific biomedical task. This technique can often be more precise than simply training directly on the problem, making it particularly useful in medical imaging and the analysis of genomic sequences. Although this method also benefits from scaling up the adequate sample size and data confidentiality, it is still inferior to its usage due to limitations arising from communication overhead issues and data heterogeneity in the distributed dataset. Nevertheless, it is among the most important steps that may be undertaken on the way to the scalable and safe application of AI in clinical settings.

Long-known machine learning (ML) algorithms that have been applied to clinical and biomedical prediction tasks include Decision Trees, Random Forests, Support Vector machines (SVMs), and k-nearest neighbors (k-NN). Decision Trees are exact, about 78.85 percent, yet significantly simplified and can be deciphered relatively easily, and thus may prove to be extremely helpful in situations where the level of transparency is of utmost importance. On the other hand, Random Forests are never less accurate than single-tree models, and are resistant to missing data and noise. In contrast, the SVMs work pretty well in high-dimensional feature spaces like genomics and transcriptomics. Lastly, k-NN can be considered relatively unsophisticated, but its similarity-based reasoning and logical and intuitive nature make it appealing and provide a variety of applications in biomedical data. These methods provide solid underpinnings as a group; however, their performance can often depend on feature engineering.

5. APPLICATION AREAS AND CASE STUDIES

5.1 Categorization of Variants and Sickness Diagnosis

Several ML-based variant pathogenicity prediction systems have been created with diverse algorithms such as Random Forest, CNN, ensemble methods, and transformer methods. These tools take huge training sets of hundreds of thousands and millions of variants and deliver high performance with moderate to high clinical adoption rates.

With NEDAMSS, two broad categories of truncating variants are of interest:

- Pathogenic, with high confidence scores, and sensitive to obvious pathogenic variants.
- In contrast to Missense variants, it must be functionally validated.

Table **5.1** depicts various tools used for predicting the probability of an infectious agent or pathogen causing a disease in a host. Each tool is mapped with the algorithm used in it, along with its demand in clinical adoption.

TABLE - 5.1: ML - Based Variant Pathogenicity Prediction Tools

Tool	Algorithm	Clinical Adoption
ClinVar RF	Random Forest	Moderate
DeepVariant	CNN	High
Primate AI	CNN + Evolution	Low
CADD	SVM + Features	High
REVEL	Ensemble	High
Alpha Missense	Transformer	Emerging

5.2 Prognosis and progression of the disease

Several methods have been developed to predict the progression of diseases, including LSTM with Clinical data, CNN with Brain MRI, Random Forest with Multi-Omics data, and Transformer with Clinical Notes. They operate on various types of data and at scales ranging from months to years, with good to high accuracy, and can be applied with varying clinical utility.

TABLE-5.2: ML/DL Approaches for Disease Progression Prediction

Approach	Data Type	Time Horizon	Clinical Utility
LSTM + Clinical	EHR, Labs	1-5 years	High
CNN + Brain MRI	Neuroimaging	6 months-2 years	High
Random Forest + Multi - Omics	Genomics, Proteomics	2-10 years	Medium
Transformer + Clinical Notes	NLP from EHR	3 months-1 year	Medium

Predicting the progression of a disease using various ML/DL algorithms is becoming crucial for better diagnosis and disease curing. Moreover, **Table 5.2** presents the various approaches used, along with their respective accuracy and prediction duration.

5.3 Repurposing & Drug Discovery

AI drug discovery promises extremely rare diseases:

- **Network-based approaches:** Identifying disease-relevant pathways in terms of medications.
- **Matching molecular signatures:** disease signatures versus drug response signatures.
- Examples of AI-identified candidates include anti-seizure drugs, HDAC inhibitors and rapamycin analogs.

5.4 NEDAMSS Disease Case Study

Comprehensive NEDAMSS Analysis Model and Predicted Measures:

- Excellent specificity of the pathogenicity of IRF2BPL variants [9].
- Strong predictive validity regarding the future (greater than 5 years).
- The length of average time to diagnose dropped to less than two years.

5.5 Comparative Analysis

Various ML-based models, including CNN with Clinical data, Random Forest, SVM with Imaging, LSTM with EEG, and GNN with Pathways, have been applied to different ultra-rare disorders, such as NEDAMSS, Rett Syndrome, Angelman Syndrome, Dravet Syndrome, and STXBP1 Encephalopathy. These have high diagnostic sensitivity, but have various problems when compared to other methods, including a lack of longitudinal data, phenotype heterogeneity, complexity of methylation, variability of seizures, and the problems of functional interpretation.

TABLE-5.3: ML/DL Performance Comparison Across Ultra-Rare Disorders

Disorder	Gene(s)	Best ML Approach	Challenges
NEDAMSS	IRF2BPL	CNN + Clinical	Limited longitudinal data
Rett Syndrome	MECP2, CDKL5	Random Forest	Phenotypic heterogeneity
Angelman Syndrome	UBE3A	SVM + Imaging	Methylation complexity
Dravet Syndrome	SCN1A	LSTM + EEG	Seizure variability
STXBP1 Encephalopathy	STXBP1	GNN + Pathways	Functional interpretation

Many ultra-rare disorders have various approaches involved while also having their own limitations, as shown in **Table 5.3**. It also describes the accuracy of diagnosis.

6. COMPARATIVE PERFORMANCE & LIMITATIONS

6.1 Summary performance measures

The performance measurements differ in other aspects of the application. Large sample sizes in Variant Pathogenicity prediction result in high accuracy, sensitivity, specificity, F1 scores, and AUC values. Syndrome Recognition on moderate sample sizes has good to high performance across measures. Disease Progression using low samples shows moderate to good results. Drug Repurposing and variable sample sizes demonstrate high performance on the metrics, and Neuroimaging Analysis and moderate sample sizes demonstrate good to high performance.

Performance Trends:

- Traditional machine learning is typically better with smaller datasets (less than several hundred samples).
- Hybrid methods are best when using medium sized data sets (hundreds to thousands of samples).
- Deep learning is the only approach that works with large datasets (thousands of samples).
- Multi-modal data: The benefits of deep learning are obvious.

6.2 Limitations and Challenges

Data Sparsity Challenges:

- **Small sample sizes:** Depending on how rare the condition is, there are frequently hundreds of patients worldwide who are characterized with that condition.
- Geographic clustering variation expectorated.

- A few of the mitigation methods are: transfer learning, data augmentation, and few-shot learning.

Overfitting symptoms are:

- Failure to generalize single site studies.
- The problem of validation in the time-space.
- Multisystem validation.

6.3 Interpretability Issues

- Understanding diagnostic thinking.
- Picking out the importance of a feature.
- Uncertainty measure and confidence measure.
- Some of the solutions include attention mechanisms, SHAP, LIME, and human-in-the-loop systems.

7. PRIVACY, SECURITY, AND ETHICAL CHALLENGES

7.1 Privacy Concerns

However, the genetic information at the center of research on rare diseases has a distinct privacy problem that is much larger than concerns in the context of conventional medical information protection. Genetic variations may be naturally detectable, posing threats to the individual patient and to relatives who are genetically related. These concerns are exacerbated by the fact that the relevance of genomic data available today may have unimaginable consequences in some decades down the line.

A technical protection measure includes:

- Advanced de-identification techniques.
- Differential privacy protocols that append judiciously tuned noise to sets of data.
- Homomorphic approaches to crypto which allow computation on ciphered data.

7.2 Security Risks

- The security risk associated with data on rare diseases not only includes threats from opportunistic cybercriminals but also includes threats from more advanced state sponsors. Ransomware attacks and data breaches pose a direct threat to patient privacy and the continuity of research, and insider threats can already exploit atypical access privileges to gain unauthorized access to health data.
- Multilayered security should involve encryption of resources both at rest and in transit, fine-grained controls that enforce principles of least privilege, and effective network security designs. Routine security auditing and updating of incident response plans are in place to ensure that countermeasure actions remain current with new attack vectors.

7.3 Bias and Fairness

The causes of bias in rare disease AI systems are similar to the existing healthcare disparities, but come with specific difficulties associated with a global distribution of rare diseases:

- Geographic bias - emerges from the concentration of research activities in high-income countries.
- Socioeconomic differences - have an impact on genetic testing and specialized care access.

This is because genetic variants potentially relevant to pathogenicity in one population group might not be the same in the other population group. Reduction measures involve holistic interventions (such as diverse recruitment campaigns that proactively aim to represent underrepresented groups), inclusive study design (reflecting the requirements of various communities), and algorithmic fairness measures that have the potential to identify and remediate systematic biases in prediction.

7.4 Ethical Concerns

The tension between need and autonomy that could have ensued in relation to whether autonomy should be respected or not in rare disease research amplifies the complexity of informed consent. With rare diagnoses, patients and families are pressured to engage in research that could potentially be beneficial to their situation, which may undermine the voluntary aspect of consent.

Key ethical issues include:

- Informed consent complexity in genomic research contexts.
- Ethics in the field of data sharing and the provision of rights to all or an individual.
- Regulatory compliance with GDPR, HIPAA, and specialized policies like Orphanet guidelines.

8. FUTURE SCOPE

8.1 Innovations in Technology

One of the most promising areas for the development of AI for rare diseases is transfer learning. Researchers will be able to use data from model organisms and related data sources to enhance predictions for rare human diseases, thanks to developments in cross-species and cross-modal transfer learning.

Important advancements consist of:

- Meta-learning strategies for few-shot learning situations that are frequently encountered in rare illness settings
- Domain adaptation strategies to ensure AI models can serve a variety of patient groups and close population gaps
- Attention-based mechanisms in multimodal fusion strategies, such as early, late, and intermediate fusion approaches.

8.2 Teamwork Methodologies

Initiatives for international data sharing that can overcome the basic drawback of small sample sizes will become increasingly important in the future of rare disease research. Standardized protocols and incentive alignment mechanisms are necessary for federated learning networks to succeed, but they offer potential solutions for protecting privacy while facilitating collaborative model creation. International collaboration, which is crucial for rare illness research, is hampered by disparate jurisdictions' differing standards for data sharing and protection, making regulatory harmonization a significant problem.

8.3 Accurate Medical Care

The final goal of AI-based applications in rare disease is custom-made pipeline to therapy:

- Genomic-guided therapy selection is the use of genetic profiles to pair patients to the most appropriate and effective medicines.
- Pharmacogenomics optimization: identification of the optimal dosage and the prediction of adverse effects.
- Adaptive clinical trial designs, early termination of ineffective treatments and efficient assignment of patients.
- The multi-omics fusion methods that bio-sensitive proteomic, metabolomic, transcriptomic, and genomic data remain useful in the process of finding biomarkers. Although wearable technology can offer continuous monitoring capabilities for rare diseases due to the digital biomarkers it generates, liquid biopsies could provide promising non-invasive monitoring options.

8.4 Integration of Digital Health

- The integration of AI and IoT technology opens possibilities that have never been imagined: continuous monitoring of patients with rare diseases. As much as smart home sensors have the capacity to assess altered behavior and functional efficacies, wearable technology may provide complete physiological data that captures disease patterns.
- Digital therapeutics is a new subcategory of therapies in which individualized care is provided through digital channels by using artificial intelligence (AI) [10]. Virtual reality applications can have potential in symptom management and rehabilitation, and the development of appropriate regulatory channels remains a significant challenge.

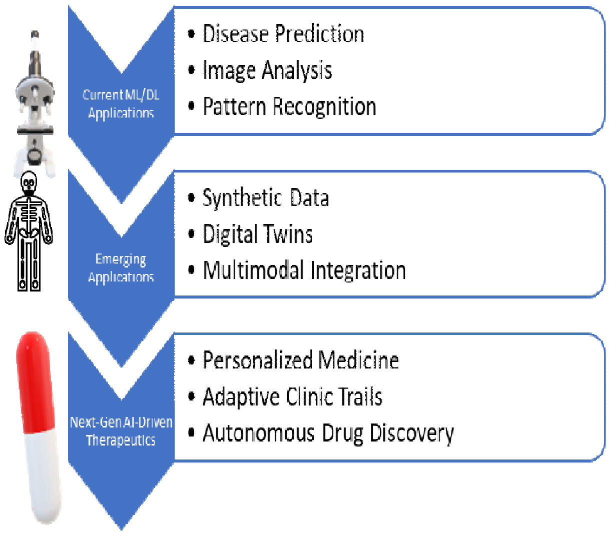

FIGURE - 8.1: Future Scope in AI Therapeutics

8.5 Treatments of the Future

With artificial intelligence under control and the integration of gene editing technology, rare genetic diseases can now be treated more efficiently. The future scope of the treatment process is depicted in **Figure 8.1**. AI-based CRISPR may be optimized to increase the accuracy and effectiveness of the gene-editing process, and predictive systems might be used to detect and manage undesired outcomes.

Some of the new treatment methods include:

- Antisense oligonucleotides generated by AI and specific to RNA sequences.
- Improved stability by machine learning-assisted protein engineering.
- AI regenerative medicine uses to guide tissue engineering and ensure full reprogramming of cells.

9. CONCLUSION

Key Findings: This systematic review has not only highlighted the significant issues at the interface between diagnosing ultra-rare genetic disorders and the transformative opportunities of ML/DL methods. We have reviewed multiple studies and found that, in various areas, there were improvements in performance, with diagnostic accuracies ranging in high ranges. The most commonly used architecture is CNNs (the majority of studies), but new specialized methods, such as transfer learning and few-shot learning, have become important in addressing the problem of data scarcity.

NEDAMSS as a Model System: NEDAMSS disease is an example of an ultra-rare disorder that has proven the potential of AI. The suggested comprehensive diagnostic pipeline will decrease the average diagnosis duration significantly and, with the help of combined analysis, will reach high accuracy rates. The longitudinal ML models can be tested ideally on the progressive nature, and the molecular characterization of IRF2BPL can present valuable biological constraints used to develop the models.

The Path Forward: It will also need interdisciplinary collaboration, like it has never been, to combine AI, clinical genetics, molecular biology, regulatory science, and patient advocacy to succeed. Technical advancements should be based on biological knowledge, and cooperation between the world via federated learning networks is the only possible way of achieving adequate sample sizes.

The result of success will be significant changes in patient outcomes, such as fewer diagnostic delays, better prognostic accuracy, and better quality of life for patients and their families with ultra-rare genetic diseases. With collective innovation, strict validation, and patient-centric development, AI-based solutions will change how the most infrequent human diseases are diagnosed, prognosed, and treated.

10. REFERENCES

[1] Lee, J., Liu, C., Kim, J., *et al.* (2022). Deep learning for rare disease: A scoping review. *Journal of Biomedical Informatics, 135,* 104227. https://doi.org/10.1016/j.jbi.2022.104227

[2] Marcogliese, P. C., *et al.* (2022). Loss of IRF2BPL impairs neuronal maintenance through Wnt signaling. *Science Advances, 8*(2), eabl5613. https://doi.org/10.1126/sciadv.abl5613

[3] Colmenarejo, G., *et al.* (2023). A systematic review on machine learning approaches in the diagnosis and prognosis of rare genetic diseases. *Journal of Biomedical Informatics, 140,* 104808. https://doi.org/10.1016/j.jbi.2023.104808

[4] Marcogliese, P. C., Shashi, V., Spillmann, R. C., Stong, N., Rosenfeld, J. A., Koenig, M. K., Campeau, P. M. (2018). IRF2BPL is associated with

neurological phenotypes. *American Journal of Human Genetics, 103*(2), 245–260. https://doi.org/10.1016/j.ajhg.2018.07.001

[5] **Simons Searchlight**. (2025). *IRF2BPL-related syndrome*. Retrieved from https://www.simonssearchlight.org

[6] Heide, S., *et al.* (2023). IRF2BPL causes mild intellectual disability followed by a progressive neurologic course. *Neurology Genetics, 9*(5), e0096. https://doi.org/10.1212/NXG.0000000000000096

[7] Abbas, S. R., *et al.* (2025). A review on machine learning applications for rare genetic disorders. *International Journal of Molecular Sciences, 26*(14), 3864. https://doi.org/10.3390/ijms26143864

[8] Chandramohan, D., Garapati, H. N., Nangia, U., Simhadri, P. K., Lapsiwala, B., Jena, N. K., & Singh, P. (2024). Diagnostic accuracy of deep learning in detection and prognostication of renal cell carcinoma: A systematic review and meta-analysis. *Frontiers in Medicine, 11*, Article 1447057. https://doi.org/10.3389/fmed.2024.1447057

[9] **Gene Reviews**. (2024, November). *IRF2BPL-related disorder*. University of Washington, Seattle. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK564234/

[10] Wojtara, M., *et al.* (2023). Artificial intelligence in rare disease diagnosis and treatment. *Clinical and Translational Science, 16*(3), 481–495. https://doi.org/10.1111/cts.13501.