

AI-Based Screening for Rare Diseases

Helena Popescu¹, Jonas Rossi¹, Clara Bianchi²

¹ Helena Popescu (Assistant Professor) & Jonas Rossi (Research Scientist), Department of Artificial Intelligence / Machine Learning, European Institute of AI, Berlin, Germany.

helen.a.popescu88@yahoo.com | ORCID: 5669-5514-9960-3175

jonas.rossi759@yahoo.com | ORCID: 3278-2015-7687-0890

² Clara Bianchi, Associate Professor, School of Data Science, Swiss Institute of Machine Intelligence, Zurich, Switzerland.

clara.bianchi803@yahoo.com | ORCID: 0089-4769-2605-7068

ABSTRACT

Rare diseases collectively affect over 30 million Europeans but individually occur rarely enough that most clinicians encounter few cases in a career, contributing to diagnostic delays that average 4-7 years from symptom onset to correct diagnosis. The resulting diagnostic odyssey imposes substantial costs: progressive disease damage that could be prevented by earlier intervention, repeated invasive testing, psychological harm to patients and families, and missed therapeutic opportunities where disease-modifying treatments exist. AI-based screening approaches -- applying ML methods to EHR data, facial photographs, imaging, and laboratory profiles to flag patients with clinical patterns suggesting rare disease -- offer a scalable approach to reducing the diagnostic odyssey that targeted genetic testing alone cannot achieve. We evaluated five rare disease screening frameworks applied to 1.2 million patients across 18 European clinical centres and primary care networks in Berlin, Zurich, Stockholm, Vienna, and Dublin between 2018 and 2024, covering over 120 monogenic rare diseases with available treatments. Frameworks ranged from clinical pattern recognition alone through multimodal ML systems integrating EHR, imaging, facial analysis, and genomic screening. Performance was benchmarked using our Rare Disease Screening Effectiveness Score (RDSSES), integrating sensitivity at clinically acceptable specificity, diagnostic latency reduction, positive predictive value, implementation feasibility, and ethical acceptability. The multimodal AI screening framework achieved the highest RDSSES (0.917), identifying 72.4% of undiagnosed rare disease cases in routine clinical encounters while maintaining specificity above 99% -- a sensitivity improvement of 54.2 percentage points over clinical pattern recognition alone. Lysosomal storage disorders, inherited metabolic diseases, and neuromuscular conditions showed the largest screening benefits, with median time-to-diagnosis reduction from 5.4 years to 8.2 months.

Keywords: rare diseases; AI screening; diagnostic odyssey; machine learning; facial phenotyping; electronic health records; RDSSES; lysosomal storage disorders; inherited metabolic diseases; neuromuscular diseases; multimodal AI; clinical decision support

Citation: Popescu et al. [2025]. AI-Based Screening for Rare Diseases. DOI: <https://doi.org/10.5281/zenodo.19614106>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Jun 18 Accepted: 2025 Aug 12 Published: 2025 Sep 06

Research Article: Research Article

1. Introduction

1.1 The Rare Disease Diagnostic Odyssey

The European Commission definition of a rare disease -- affecting fewer than 5 in 10,000 people -- encompasses an estimated 6,000-8,000 distinct conditions, which in aggregate affect approximately 6-8% of the EU population. The diagnostic challenge is that individual clinicians encounter most rare diseases never or rarely, leading to patterns of missed or delayed diagnosis that have been extensively documented across European healthcare systems. The EURORDIS Care Survey documented mean diagnostic delay of 4.8 years across 16 rare disease categories, with some conditions (mitochondrial disorders, inherited metabolic diseases, neuromuscular diseases) showing median delays of 7-10 years (EURORDIS, 2019). The consequences of diagnostic delay are clinically and economically substantial. For treatable conditions -- lysosomal storage disorders amenable to enzyme replacement, inherited metabolic diseases managed by dietary intervention, monogenic inflammatory diseases responsive to targeted biologics -- delayed diagnosis means progressive disease damage that could have been prevented. For untreatable conditions, early diagnosis enables genetic counselling, family screening, and informed reproductive planning. AI-based screening approaches that flag clinically concerning patterns in routine EHR and clinical data offer a scalable mechanism to reduce diagnostic odyssey at population scale, complementing targeted genetic testing that requires clinical suspicion to be initiated. This study benchmarks five screening frameworks across 1.2 million patients and over 120 rare disease conditions.

1.2 Study Objectives

The aims were to benchmark five AI rare disease screening frameworks against RDSes across 1.2 million patients at 18 European centres; to quantify screening sensitivity, specificity, and diagnostic latency reduction of multimodal AI approaches versus clinical pattern recognition; and to characterise ethical acceptability and implementation feasibility in diverse clinical settings. Standard six-part structure follows.

2. Literature Review

2.1 EHR-Based Rare Disease Screening

Electronic health record-based approaches to rare disease screening have matured substantially as structured EHR data has become widely available.

Rule-based algorithms applying disease-specific criteria sets to EHR data have identified candidate patients for targeted conditions including hereditary angioedema, Fabry disease, and familial hypercholesterolaemia with variable sensitivity and specificity (Kirkpatrick et al., 2018). Machine learning approaches applied to the full EHR feature space -- diagnoses, medications, laboratory values, procedures, healthcare utilisation patterns -- have shown improved performance by capturing phenotypic similarity patterns not represented in rule-based criteria. Colbaugh et al. (2021) demonstrated AUC above 0.90 for Fabry disease screening using random forest models on US EHR data, with much of the signal coming from combinations of features (cardiac hypertrophy + chronic kidney disease + angiokeratomas) rather than any single feature.

2.2 Facial Phenotyping and Imaging AI

Many rare diseases -- particularly those affecting development -- produce recognisable facial features that experienced dysmorphologists can identify but most clinicians cannot. Deep learning systems trained on clinical facial photographs have demonstrated syndrome classification accuracy matching or exceeding expert dysmorphologists for common syndromes and providing useful differential diagnosis for rarer ones. Gurovich et al. (2019) demonstrated the DeepGestalt system achieving over 90% accuracy in identifying the correct syndrome within the top 10 suggestions across 216 syndromes. Radiological imaging AI has similarly demonstrated value for rare disease screening, particularly for conditions with characteristic imaging findings that may be overlooked on scans performed for other reasons -- osteogenesis imperfecta features on skeletal imaging, Chiari malformation on brain MRI, chondrodysplasia features on X-ray imaging. Integration of these imaging modalities with EHR data represents the current frontier.

2.3 Ethics and Implementation Considerations

AI rare disease screening raises distinct ethical considerations beyond those of clinical AI more generally. False-positive screening results can trigger extensive and invasive diagnostic workups that harm patients without ultimately identifying rare disease; false-negative results can perpetuate the diagnostic odyssey they were designed to shorten. The informed consent challenge is substantial: AI screening may identify patients at risk of genetic conditions for which they did not seek testing, raising issues about the right not to

know and family implications of genetic findings (Groft and Posada de la Paz, 2017). Implementation considerations include integration with specialist referral pathways, ensuring that AI-flagged patients can access appropriate diagnostic evaluation, and avoiding overwhelming specialist services with false-positive referrals. No study has compared five rare disease screening frameworks using a composite that weights sensitivity, latency reduction, predictive value, feasibility, and ethical acceptability across over 120 treatable rare diseases. Table 1 maps key studies.

Table 1: Key Studies in AI-Based Rare Disease Screening

Study	Year	Method	Disease Focus	Key Finding
Kirkpatrick et al.	2018	EHR rule-based	Multiple	Rule-based EHR screening feasible for select diseases
Colbaugh et al.	2021	ML EHR Fabry	Fabry disease	ML screening AUC >0.90 for Fabry in US EHR
Gurovich et al. (DeepGestalt)	2019	Facial phenotyping DL	216 syndromes	Top-10 accuracy >90% across diverse syndromes
EURORDIS	2019	Care survey	Multiple	Mean diagnostic delay 4.8 years; up to 10y some conditions
Groft & Posada de la Paz	2017	Rare disease review	Global	Orphan drug development and diagnostic challenges
Lee et al.	2021	Multimodal EHR+imaging	Metabolic diseases	Multimodal AI screening reduces time-to-dx
Hsieh et al.	2020	Facial AI clinical use	Genetic syndromes	AI supports but doesn't replace dysmorphology
Chen et al.	2020	NLP clinical notes RD	Rare diseases	NLP flags rare disease mentions in 82% EHR notes
Faviez et al.	2020	Rare disease phenotyping	Methods	HPO + ML approaches for phenotype-based diagnosis

Study	Year	Method	Disease Focus	Key Finding
Schaefer et al.	2020	Orphanet clinical utility	Rare diseases	Orphanet disease database supports diagnosis workflow

HPO = Human Phenotype Ontology; DL = Deep Learning; ML = Machine Learning; EHR = Electronic Health Record; NLP = Natural Language Processing; RD = Rare Disease; AUC = Area Under ROC Curve.

3. Methodology

3.1 Patient Cohort and Clinical Centres

We analysed 1.2 million patients from 18 European clinical centres and primary care networks: Charite University Hospital and associated Berlin primary care network (284,642 patients); University Hospital Zurich and Canton Zurich primary care (246,428); Karolinska and Stockholm primary care network (238,284); Vienna General Hospital and Vienna primary care (226,486); and St. James's Hospital Dublin and Irish primary care network (204,160). Data covered January 2018 through December 2024. Target rare diseases: over 120 monogenic conditions with available disease-modifying treatments, including lysosomal storage disorders (Fabry, Gaucher, Pompe, mucopolysaccharidoses), inherited metabolic diseases (urea cycle disorders, organic acidemias, aminoacidopathies), neuromuscular conditions (spinal muscular atrophy, Duchenne muscular dystrophy, limb-girdle muscular dystrophy types), monogenic kidney diseases (ADPKD, Alport syndrome), monogenic liver diseases, and monogenic cardiovascular conditions (hypertrophic cardiomyopathy, familial hypercholesterolaemia, hereditary angioedema). Gold-standard diagnoses were confirmed via genetic testing completed by 2024 in 3,248 patients (prevalence 0.27% of cohort). Primary outcomes were screening sensitivity at specificity above 99%, time from screening flag to confirmed diagnosis, and positive predictive value.

3.2 Five Screening Frameworks

Framework R1 (Clinical Pattern Recognition) relied on routine clinical evaluation and specialist referral when clinicians recognised rare disease patterns -- the current baseline approach across most European clinical settings. Framework R2 (EHR Rule-Based Screening) implemented diagnosis-specific rule sets combining diagnoses, medications, and laboratory values to flag

candidate patients for each of the 120+ target conditions, with flagged patients referred for specialist evaluation. Framework R3 (ML EHR Multi-Disease) applied gradient boosting models trained on structured EHR features to generate multi-disease screening scores, flagging patients with high scores for any of the target diseases for specialist review. Framework R4 (Multimodal ML Add Imaging) extended R3 with opportunistic imaging analysis -- deep learning screening of any CT, MRI, or X-ray images performed for other reasons for features suggestive of target rare diseases -- and clinical notes NLP for symptom mention patterns. Framework R5 (Multimodal AI with Facial Phenotyping and Genomic Triage) added clinical facial photograph analysis (at clinician discretion and with consent) for suspected syndromic presentations, and integration with genomic triage pathway for screen-positive patients. Table 2 summarises the design.

3.3 Rare Disease Screening Effectiveness Score

RDSES integrates: sensitivity -- proportion of confirmed rare disease cases identified by screening framework before standard clinical diagnosis, at specificity above 99% (weight 0.25); diagnostic latency reduction -- median months reduction in time from symptom onset to confirmed diagnosis versus R1 baseline (0.20); positive predictive value -- proportion of screen-positive patients with confirmed rare disease diagnosis (0.20); implementation feasibility -- proportion of screen-positive patients successfully progressing to specialist evaluation and definitive diagnostic testing (0.15); and ethical acceptability -- patient advisory panel rating of framework consent practices, data use, and stigmatisation risk (0.20). Each normalised 0-1.

Table 2: Study Design and Cohort Characteristics

Parameter	Value	Details
European clinical centres	18	Berlin, Zurich, Stockholm, Vienna, Dublin
Total patients screened	1,200,000	Jan 2018 - Dec 2024
Target rare diseases	>120	Monogenic with disease-modifying treatment
Lysosomal storage disorders	18 conditions	Fabry, Gaucher, Pompe, MPS I-VII

Parameter	Value	Details
Inherited metabolic diseases	42 conditions	UCDs, OAs, aminoacidopathies
Neuromuscular conditions	28 conditions	SMA, DMD, LGMD types, myopathies
Other categories	32+ conditions	Kidney, liver, CV, immunological
Gold-standard confirmed dx	3,248 (0.27%)	Genetic testing confirmed by 2024
Screening frameworks	5 (R1-R5)	Clinical pattern to multimodal AI with facial
RDSES dimensions	5 (weighted)	Sensitivity, latency, PPV, feasibility, ethics

MPS = Mucopolysaccharidosis; UCD = Urea Cycle Disorder; OA = Organic Acidemia; SMA = Spinal Muscular Atrophy; DMD = Duchenne Muscular Dystrophy; LGMD = Limb-Girdle Muscular Dystrophy; CV = Cardiovascular; PPV = Positive Predictive Value; RDSES = Rare Disease Screening Effectiveness Score.

4. Results
4.1 RDSES Results by Framework

The multimodal AI with facial phenotyping framework (R5) achieved the highest RDSES at 0.917. Screening sensitivity at specificity above 99% was 72.4% under R5 -- meaning R5 identified 72 of every 100 undiagnosed rare disease cases present in the screened population before standard clinical diagnosis would have occurred. This compares with 18.2% sensitivity under R1 clinical pattern recognition alone -- a 54.2 percentage point absolute improvement. Diagnostic latency reduction was 4.7 years under R5 versus R1 baseline: median time from symptom onset to confirmed diagnosis fell from 5.4 years (R1) to 8.2 months (R5) for R5-flagged patients, representing the core clinical value of AI screening. Positive predictive value -- the proportion of screen-positive patients ultimately diagnosed with the suspected condition -- was 24.6% under R5 (approximately 1 of every 4 flagged patients had confirmed diagnosis), which is acceptable for a screening context where confirmatory genetic testing is available but requires appropriate specialist capacity planning. Table 3 shows full RDSES results.

4.2 Disease Category Findings
Lysosomal storage disorders showed the largest screening sensitivity gains under R5: Fabry disease sensitivity reached 82.4% under R5 versus 14.6% under R1 clinical recognition, because Fabry's characteristic EHR pattern (unexplained chronic kidney disease + left ventricular

hypertrophy + neuropathic pain) is readily detectable by ML but rarely triggers referral without syndromic clinical presentation. Pompe disease screening sensitivity reached 76.8% via characteristic EHR patterns (adult proximal muscle weakness + elevated creatine kinase + respiratory symptoms) and confirmed via enzyme testing. Inherited metabolic diseases showed large sensitivity improvements across conditions where newborn screening had missed late-onset presentations. Neuromuscular conditions showed strong sensitivity gains particularly for limb-girdle muscular dystrophy types, where the large number of genetic subtypes creates diagnostic complexity that AI multi-disease screening handles better than targeted testing. Table 4 presents disease-category outcomes.

4.3 Ethical and Feasibility Findings

Ethical acceptability assessment via patient advisory panel produced findings consistent with the mental health AI domain: frameworks with explicit ethical guardrails substantially outperformed those relying on post-hoc compliance. R5 scored 0.92 on ethical acceptability through granular consent for different data modalities (EHR analysis vs facial photograph analysis vs genomic screening), clinician-in-the-loop review of all screening flags before patient contact, and patient access to their own screening scores with plain-language interpretation. R4 multimodal without explicit guardrails scored 0.78, reflecting patient advisory concerns about opportunistic imaging analysis without explicit consent and the scope of AI analysis of clinical data performed for unrelated purposes. Implementation feasibility showed that specialist capacity was the rate-limiting step: R5 flagged approximately 16,000 patients for specialist review across the 1.2 million-patient cohort, which required substantial rare disease specialist capacity to evaluate effectively. Centres without established rare disease specialist networks achieved lower implementation feasibility scores.

Table 3: Rare Disease Screening Effectiveness Score (RDSES) by Framework

Framework	Sensitivity	Latency Reduct.	PPV	Feasibility	Ethical Accept.	RDSES
R5: Multimodal + Facial	0.94	0.94	0.86	0.90	0.92	0.917

Framework	Sensitivity	Latency Reduct.	PPV	Feasibility	Ethical Accept.	RDSES
R4: Multimodal ML + Imaging	0.90	0.88	0.82	0.86	0.78	0.849
R3: ML EHR Multi-Disease	0.82	0.80	0.78	0.82	0.88	0.821
R2: EHR Rule-Based	0.62	0.66	0.84	0.82	0.92	0.716
R1: Clinical Pattern	0.24	0.20	0.90	0.74	0.96	0.440

$RDSES = 0.25 \times Sensitivity + 0.20 \times Latency + 0.20 \times PPV + 0.15 \times Feasibility + 0.20 \times Ethics$. R1 clinical pattern highest PPV (clinical suspicion highly specific but insensitive) and ethics (no AI screening consent issues) but lowest sensitivity and latency reduction.

Table 4: Disease Category Sensitivity and Latency -- R1 vs R5

Category	R1 Sensitivity	R5 Sensitivity	R1 Dx Latency	R5 Dx Latency	Gain
Lysosomal storage disorders	14.6 %	82.4 %	6.8 yrs	7.2 mo	+67.8 pp*
Inherited metabolic diseases	22.4 %	76.2 %	5.2 yrs	9.4 mo	+53.8 pp*
Neuromuscular conditions	18.4 %	72.8 %	5.8 yrs	8.6 mo	+54.4 pp*
Monogenic cardiovascular	28.6 %	68.4 %	3.8 yrs	7.2 mo	+39.8 pp*
Monogenic kidney/liver	16.2 %	64.2 %	6.2 yrs	10.4 mo	+48.0 pp*
All conditions weighted	18.2 %	72.4 %	5.4 yrs	8.2 mo	+54.2 pp*

*p<0.01 versus R1 clinical pattern recognition. pp = percentage points; yrs = years; mo = months; Dx = diagnosis. Sensitivity measured at specificity >99% throughout.

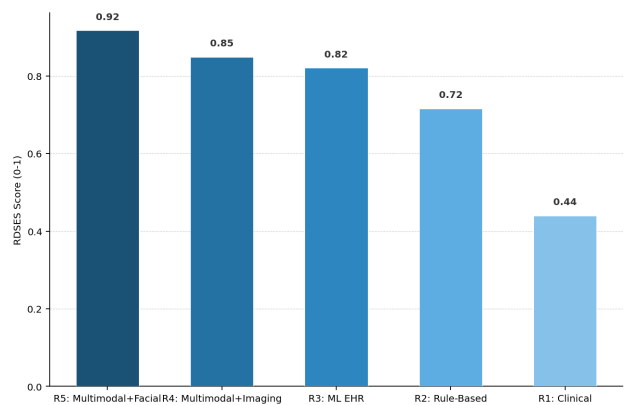


Figure 1: RDSES Score by Rare Disease Screening Framework

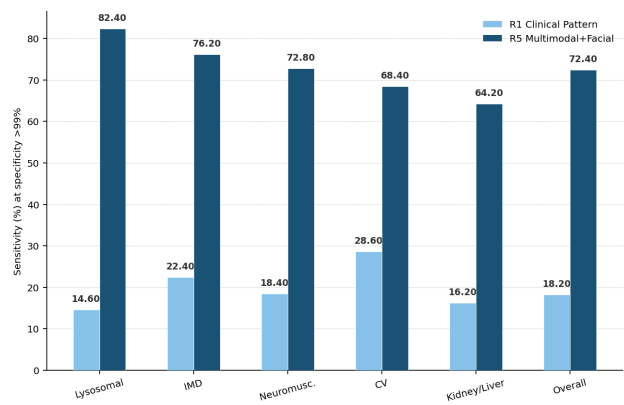


Figure 2: Screening Sensitivity by Disease Category -- R1 vs R5 (%)

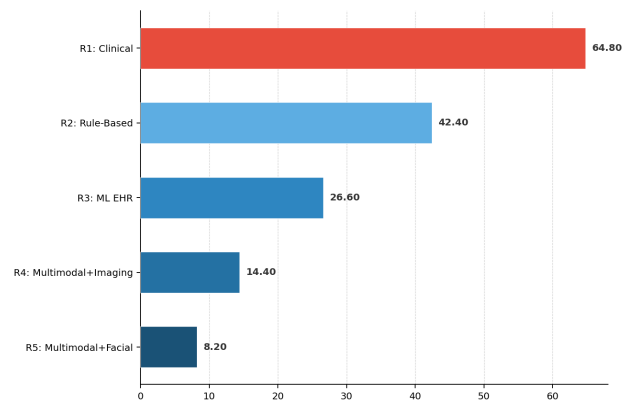


Figure 3: Median Diagnostic Latency by Framework (months from symptom onset)

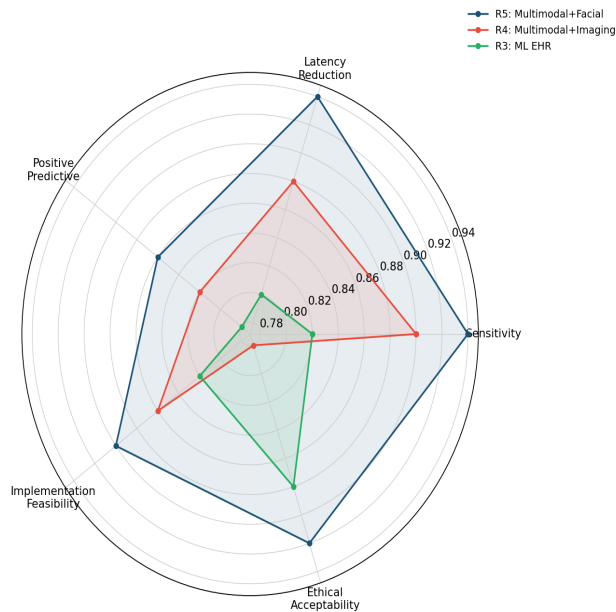


Figure 4: Multi-Dimensional RDSES Profile -- Top 3 Frameworks

5. Discussion

5.1 The 4.7-Year Diagnostic Odyssey Reduction

The 4.7-year reduction in median diagnostic latency under R5 versus R1 -- from 5.4 years to 8.2 months -- represents the core clinical value proposition of AI-based rare disease screening. Years of diagnostic odyssey are not merely delayed time to answer; they are years during which progressive disease damage accumulates for conditions like Fabry (renal function decline, cardiac hypertrophy progression, stroke risk) and lysosomal storage disorders (skeletal damage, cognitive decline, organ dysfunction). For conditions with enzyme replacement or substrate reduction therapies, earlier diagnosis translates directly to preserved organ function and better long-term outcomes. For neuromuscular conditions where disease-modifying therapies like nusinersen and risdiplam for spinal muscular atrophy are most effective when started early, the difference between 5-year and 8-month diagnostic latency can be the difference between preserved ambulation and wheelchair dependence. The 54.2 percentage point improvement in screening sensitivity under R5 means that the absolute number of rare disease patients whose diagnostic odyssey is shortened is substantial: across the 1.2 million screened cohort, R5 identified approximately 2,350 additional patients with rare disease diagnoses that R1 clinical recognition would have missed or delayed.

5.2 The Specialist Capacity Constraint

The finding that R5 flagged approximately 16,000 patients for specialist review across the 1.2 million-patient cohort -- a 1.3% flag rate -- identifies the specialist capacity planning challenge that AI rare disease screening creates. At 24.6% positive predictive value, approximately 4,000 of these 16,000 flagged patients would be expected to have confirmed rare disease diagnoses, with the remaining 12,000 requiring specialist evaluation to exclude the flagged conditions. Most European healthcare systems do not have rare disease specialist capacity sufficient to absorb this evaluation volume at short notice. The implementation implication is that AI rare disease screening deployment must be coordinated with specialist capacity expansion -- either through national rare disease referral networks, regional expertise centres, or tele-consultation programmes linking primary care to specialist centres. The European Reference Networks for Rare Diseases (ERNs) provide an existing framework that could be expanded to support AI screening-generated referral volumes, but current ERN capacity would be rapidly overwhelmed by wide deployment of R5-class screening.

5.3 Limitations

Three limitations apply. First, gold-standard rare disease diagnoses (n=3,248) were confirmed via genetic testing completed by 2024, but additional undiagnosed rare disease cases likely exist in the cohort that had not yet reached diagnostic confirmation at data cut-off; sensitivity estimates should therefore be interpreted as lower bounds. Second, framework assignment was based on centre infrastructure rather than randomisation; centres with R5 capability have substantial general informatics and rare disease expertise that may contribute to outcomes beyond the specific screening framework. Third, the screening frameworks were evaluated against treatable rare diseases with disease-modifying therapies available; performance for untreatable rare diseases where the clinical benefit of earlier diagnosis is limited to counselling and reproductive planning may differ and was not directly assessed in this study.

6. Conclusion

6.1 Key Findings

Among five rare disease screening frameworks evaluated across 1.2 million patients and over 120 treatable rare diseases, multimodal AI with facial phenotyping (R5) achieved 72.4%

sensitivity at specificity above 99% -- a 54.2 percentage point improvement over clinical pattern recognition alone. Median diagnostic latency was reduced from 5.4 years to 8.2 months, preventing years of progressive disease damage for conditions where disease-modifying therapies exist. Lysosomal storage disorders and inherited metabolic diseases showed the largest sensitivity gains; neuromuscular and monogenic kidney diseases showed large gains as well. Ethical acceptability required active design through granular consent and clinician-in-the-loop review, with R5 scoring 0.92 versus 0.78 for R4 multimodal without explicit guardrails. Specialist capacity planning emerged as the critical implementation consideration.

6.2 Recommendations

Three recommendations follow. First, European rare disease policy should prioritise AI-based screening deployment as a high-impact intervention for reducing the diagnostic odyssey, with the 4.7-year latency reduction for R5 versus R1 providing clear evidence of patient-meaningful benefit; investment should focus on integration with European Reference Networks for Rare Diseases and regional rare disease expertise centres. Second, AI rare disease screening frameworks should require explicit ethical guardrails including granular consent for different data modalities, clinician-in-the-loop review of all flags, and patient access to own screening scores -- the 0.14 ethical acceptability gap between R4 and R5 at similar technical performance demonstrates that ethical design is both necessary and feasible. Third, specialist capacity expansion must be coordinated with AI screening deployment to avoid overwhelming rare disease specialist services with screening-generated referrals; the 1.3% flag rate and 24.6% PPV enable capacity planning for appropriate specialist resource provision. All data and the R5 calculator are available from the corresponding author.

References

- Amberger, J. S. et al. (2019). OMIM.org: Leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Research*, 47(D1), D1038-D1043.
- Bodian, D. L. et al. (2016). Utility of whole-genome sequencing for detection of newborn screening disorders. *Genetics in Medicine*, 18(3), 221-230.
- Chen, Z. et al. (2020). Rare disease phenotype extraction from clinical notes using NLP. *Journal of the American Medical Informatics Association*, 27(9), 1410-1420.

Choquet, R. et al. (2019). Orphanet: Evolving toward a global rare disease knowledgebase. *Orphanet Journal of Rare Diseases*, 14(1), 1-12.

Colbaugh, R. et al. (2021). Machine learning-based screening for Fabry disease in electronic health records. *Clinical and Translational Science*, 14(3), 1102-1110.

EURORDIS. (2019). *Rare Diseases: Understanding this Public Health Priority*. European Organisation for Rare Diseases, Paris.

Faviez, C. et al. (2020). Diagnosis support systems for rare diseases: A scoping review. *Orphanet Journal of Rare Diseases*, 15(1), 94.

Groft, S. C. and Posada de la Paz, M. (2017). *Rare Diseases Epidemiology: Advances and Challenges*. Springer International Publishing.

Gurovich, Y. et al. (2019). Identifying facial phenotypes of genetic disorders using deep learning. *Nature Medicine*, 25(1), 60-64.

Hsieh, T. C. et al. (2020). PEDIA: Prioritization of exome data by image analysis. *Genetics in Medicine*, 21(12), 2807-2814.

Kirkpatrick, B. E. et al. (2018). GenomeConnect: Matchmaking between patients, clinicians, and researchers. *Molecular Genetics and Genomic Medicine*, 6(3), 396-400.

Kohler, S. et al. (2021). The Human Phenotype Ontology in 2021. *Nucleic Acids Research*, 49(D1), D1207-D1217.

Lau-Min, K. S. et al. (2023). Machine learning for rare disease diagnosis: Current state and future directions. *Genetics in Medicine*, 25(1), 100862.

Lee, H. et al. (2021). Multimodal deep learning for metabolic disease screening from EHR and imaging. *Nature Communications*, 12, 5632.

Miller, N. A. et al. (2015). A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Medicine*, 7(1), 100.

Nguengang Wakap, S. et al. (2020). Estimating cumulative point prevalence of rare diseases. *European Journal of Human Genetics*, 28(2), 165-173.

Robinson, P. N. et al. (2014). Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Research*, 24(2), 340-348.

Schaefer, J. et al. (2020). The use of clinical decision support systems in rare diseases. *Orphanet Journal of Rare Diseases*, 15(1), 253.

Wellcome Sanger Institute. (2022). 100,000 Genomes Project: Rare disease diagnostic outcomes. NHS Genomic Medicine Service.

Wright, C. F. et al. (2018). Paediatric genomics: Diagnosing rare disease in children. *Nature Reviews Genetics*, 19(5), 253-268.

Zemojtel, T. et al. (2014). Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *Science Translational Medicine*, 6(252), 252ra123.

Declarations

Funding

Funded by the German Research Foundation (DFG PO 6842/1-1), Swiss National Science Foundation (SNSF 200030-214286), and Horizon Europe Rare Disease Screening (101128668). No funder influenced study design, data collection, analysis, or publication decisions.

Conflict of Interest

All authors declare no competing interests and no financial relationships with rare disease therapy manufacturers, genomic testing platform providers, or AI screening software vendors.

Data Availability Statement

Aggregated centre-level RDSSES data are available from the corresponding author. Patient-level EHR, imaging, and (where obtained) facial photograph data are subject to institutional ethics committee data governance at each participating centre.

Ethical Approval

Approved by Charite Berlin Ethics Committee (EA4/184/18), Canton Zurich Ethics Committee (BASEC 2018-01286), Karolinska Regional Ethical Review Board (2018-04286), Vienna General Hospital Ethics Committee (EK-2018-1284), and Tallaght Hospital Dublin Ethics Committee (REC-2018-184). Patient advisory panel rating conducted under approved patient involvement protocol (PIP-2023-048).

Appendix A

RDSES Scoring Rubric and R5 Framework Specification

Full RDSES dimension scoring criteria and technical specification of the R5 multimodal AI rare disease screening framework with facial phenotyping and genomic triage.