Volume 4, Issue 3, 2026

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

EXPLAINABLE MULTI-OMICS AI FOR PRECISION RISK STRATIFICATION OF INHERITED DISORDERS IN PAKISTANI FAMILIES

Dr. Habib Khan[*1], Maida Tanveer[2]

[*1]*Associate Professor, Department of Biotechnology, Abasyn University, Peshawar*
[2]*Student, Department of Biotechnology, Quaid-i-Azam University, Islamabad*

[*1]habib.khan@abasyn.edu.pk, [2]itsmaidatanveer@gmail.com

DOI:https://doi.org/10.5281/zenodo.22843154

Keywords
Explainable Artificial Intelligence; Multi-Omics; Inherited Disorders; Precision Medicine; Risk Stratification; Pakistani Families.

Article History
Received: 19 January 2026
Accepted: 04 March 2026
Published: 18 March 2026

Copyright @Author
Corresponding Author: *
Dr. Habib Khan

Abstract

Inherited disorders present substantial challenges for accurate diagnosis and individualized risk assessment, particularly in genetically diverse and highly consanguineous populations. This study examined an explainable multi-omics artificial intelligence (AI) framework for precision risk stratification of inherited disorders among Pakistani families. A quantitative predictive design was employed using integrated genomic, transcriptomic, proteomic, metabolomic, familial, and clinical indicators. Multiple machine-learning algorithms were compared using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC), while SHAP-based explainable AI was applied to identify the principal contributors to individual risk predictions. The analysis indicated that Gradient Boosting achieved the strongest predictive performance (accuracy = 87.6%; AUC = .94), compared with genomic information alone (accuracy = 78.9%; AUC = .83). Integration of clinical and multi-omics information increased the AUC from .83 to .94. Pathogenic genomic variants, multi-omics risk scores, clinical phenotype severity, family history, and consanguinity emerged as major predictive features. The findings suggest that integrating heterogeneous molecular and clinical information with explainable AI can provide a more comprehensive and interpretable approach to inherited-disorder risk stratification. The framework may support precision diagnosis, genetic counselling, family screening, and individualized clinical decision-making in Pakistan. However, the reported numerical results require validation using real-world, longitudinal, and independently collected Pakistani datasets.

INTRODUCTION

Inherited disorders represent a major challenge for precision healthcare because their clinical manifestations are often heterogeneous, genetically complex, and difficult to distinguish using phenotype-based assessment alone. Advances in next-generation sequencing have substantially improved the identification of pathogenic variants, while multi-omics approaches increasingly enable integration of genomic, transcriptomic, proteomic, metabolomic, and clinical information to characterize disease mechanisms at multiple biological levels (Hong et al., 2024; Mani et al., 2025). However, translating high-dimensional molecular data into clinically interpretable risk estimates remains a major challenge.

Artificial intelligence (AI) and machine-learning techniques provide opportunities to integrate heterogeneous biological datasets, identify nonlinear relationships, detect molecular patterns, and support individualized disease-risk prediction.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Recent research indicates that machine learning is increasingly being incorporated into genome-based precision medicine for rare genetic disorders, particularly for disease classification, variant prioritization, biomarker discovery, and individualized risk assessment (Advancing genome-based precision medicine, 2025). Multi-omics integration can further improve biological characterization because individual omics layers capture different aspects of disease processes that cannot be fully represented by genomic information alone.

Pakistan provides an important context for investigating inherited disorders because consanguineous families and extensive genetic diversity create both a substantial clinical burden and valuable opportunities for identifying recessive disease variants. Recent genomic studies involving Pakistani families have identified pathogenic, likely pathogenic, and previously unreported variants associated with neurodevelopmental, neuromuscular, retinal, skeletal, and reproductive disorders (Bibi et al., 2024; Basharat et al., 2024; Kakar et al., 2024; Aleem et al., 2025). A systematic review of hereditary neurological disorders in Pakistan further reported a predominance of homozygous causative variants, highlighting the importance of genomic investigation in affected families (2025).

Despite these advances, genetic diagnosis in Pakistan remains constrained by limited access to advanced genomic testing, fragmented clinical and molecular information, inadequate population-specific reference data, and limited integration of genomic information with other biological and clinical datasets. Recent evidence from Pakistani families demonstrates that whole-exome and whole-genome sequencing can substantially improve diagnostic yield, including in families with previously undiagnosed disorders (Bibi et al., 2024; Basharat et al., 2024; Aleem et al., 2025). Nevertheless, sequencing-based diagnosis primarily identifies candidate variants; translating these findings into comprehensive, individualized risk stratification requires integration of multiple biological and clinical dimensions.

Explainable AI (XAI) offers an important mechanism for addressing the interpretability challenge. In precision medicine, clinicians need to understand not only the predicted risk but also the molecular and clinical factors contributing to that prediction. Recent literature emphasizes that trustworthy AI requires interpretable models, appropriate validation, and integration of biological knowledge rather than reliance on predictive accuracy alone (Hong et al., 2024; Explainable biology for improved therapies in precision medicine, 2024). Therefore, integrating multi-omics information with explainable AI may provide a more comprehensive framework for identifying individuals and families at elevated risk of inherited disorders.

Accordingly, this study proposes an Explainable Multi-Omics AI framework for precision risk stratification of inherited disorders in Pakistani families. The framework integrates genomic and other molecular information with clinical and familial characteristics to develop interpretable risk predictions. By emphasizing population-specific data and explainability, the study addresses the gap between genomic discovery and clinically meaningful risk stratification and contributes to the emerging field of AI-enabled precision medicine.

Problem Statement

Inherited disorders remain difficult to diagnose and manage because many conditions exhibit substantial phenotypic overlap, genetic heterogeneity, variable penetrance, and complex genotype–phenotype relationships. These challenges are particularly relevant in Pakistan, where consanguineous families provide important opportunities for identifying recessive disorders but also contribute to a substantial burden of inherited disease. Recent sequencing studies involving Pakistani families have identified numerous pathogenic and previously unreported variants, demonstrating both the genetic diversity of affected populations and the continuing diagnostic gap (Bibi et al., 2024; Basharat et al., 2024; Aleem et al., 2025).

Conventional genetic diagnosis generally emphasizes genomic variant detection and clinical. Although whole-exome and whole-genome sequencing have considerably improved diagnostic

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

capabilities, genomic data alone may not adequately capture functional consequences, biological pathways, disease mechanisms, or interactions among molecular and clinical characteristics. Multi-omics approaches provide an opportunity to address this limitation by integrating complementary biological layers. However, the high dimensionality, heterogeneity, and interdependence of multi-omics data create substantial analytical challenges that conventional statistical approaches may not adequately address. AI and machine-learning methods can process high-dimensional datasets and identify complex patterns that may be difficult to detect through conventional analytical techniques. Nevertheless, many predictive AI models function as black boxes, limiting their clinical acceptability and making it difficult for healthcare professionals to understand why a particular individual or family has been classified as high risk. Inherited-disorder prediction therefore requires not only high predictive performance but also transparent and biologically meaningful explanations.

The research gap is particularly evident in the Pakistani context. Existing studies have predominantly focused on identifying disease-causing variants within specific disorders or families through sequencing approaches. Comparatively limited research has examined an integrated framework combining multi-omics information, familial characteristics, clinical phenotypes, AI-based prediction, and explainability for individualized risk stratification. Consequently, there is a need for a population-relevant framework capable of transforming complex molecular and clinical information into interpretable risk estimates.

This study addresses this gap by developing an explainable multi-omics AI framework for precision risk stratification of inherited disorders among Pakistani families. The proposed approach seeks to integrate heterogeneous biological and clinical information, improve identification of individuals at elevated genetic risk, prioritize relevant molecular features, and provide interpretable evidence that can support genetic counselling, clinical decision-making, and family-level preventive strategies.

Research Questions

1. Which genomic, transcriptomic, proteomic, and clinical indicators are most strongly associated with inherited-disorder risk among Pakistani families?
2. To what extent can multi-omics AI models accurately stratify individuals according to inherited-disorder risk?
3. Which AI algorithm provides the most reliable risk-stratification performance?
4. How does integrating multiple omics layers improve predictive performance compared with genomic or clinical information alone?
5. How can explainable AI identify and interpret the molecular and clinical factors underlying individual risk predictions?

Research Objectives

1. To identify major genomic, multi-omics, familial, and clinical indicators associated with inherited-disorder risk in Pakistani families.
2. To develop an AI-based multi-omics model for individualized risk stratification.
3. To compare the predictive performance of alternative machine-learning algorithms using appropriate classification and discrimination metrics.
4. To determine whether multi-omics integration improves risk-stratification performance compared with single-source data.
5. To apply explainable AI techniques to identify the major molecular and clinical contributors to individual risk predictions.
6. To propose an interpretable precision-medicine framework applicable to genetic counselling and inherited-disorder risk assessment in Pakistan.

Significance of the Study

Theoretical Significance

The study contributes to precision-medicine and AI literature by integrating multi-omics analytics with explainable machine learning. It advances the conceptualization of inherited-disorder assessment from single-variant identification toward multidimensional, individualized risk stratification.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Practical Significance

The proposed framework can support clinicians and genetic counsellors by combining molecular, familial, and clinical information into interpretable risk profiles. Such information may facilitate earlier diagnostic investigation, variant prioritization, family screening, and personalized counselling.

Policy Significance

The study can inform the development of population-specific genomic medicine strategies in Pakistan. Evidence generated through explainable AI may support investment in genomic infrastructure, genetic counselling, secure health-data systems, and standardized genomic-data collection while emphasizing appropriate ethical oversight and protection of genetic privacy.

Overall, the study addresses an important translational gap between genomic discovery and clinically interpretable precision risk assessment, particularly for Pakistani families that remain underrepresented in global genomic datasets.

Literature Review

Multi-Omics and Precision Risk Stratification

Precision medicine increasingly relies on integrating heterogeneous biological information rather than interpreting individual genetic variants in isolation. Genomics identifies sequence-level variation, whereas transcriptomics, proteomics, metabolomics, and related omics layers provide complementary information about gene expression, protein activity, metabolic pathways, and functional consequences. Recent reviews argue that this multi-layer perspective is particularly valuable for rare and inherited disorders, where phenotype–genotype relationships can be highly heterogeneous (Hong et al., 2024; Mani et al., 2025). Multi-omics integration therefore provides a conceptual basis for moving from variant identification toward individualized biological risk profiling.

However, multi-omics integration is technically challenging. Different datasets differ in scale, dimensionality, missingness, measurement platforms, biological resolution, and statistical distributions. Recent work emphasizes that effective multi-omics models must address data harmonization, dimensionality reduction, batch effects, incomplete observations, and meaningful biological rather than simply concatenating datasets (Li et al., 2026). Thus, the principal research challenge is not merely acquiring multiple omics datasets but integrating them in a manner that preserves their biological meaning.

AI and Machine Learning in Inherited Disorders

AI and machine learning have increasingly been applied to rare genetic disorders for disease classification, variant prioritization, biomarker identification, diagnosis, and individualized risk assessment. Recent evidence indicates that machine-learning models can process high-dimensional genomic information and detect nonlinear relationships that may be difficult to identify using conventional statistical approaches (Advancing genome-based precision medicine, 2025). AI-based methods are particularly relevant when numerous variants interact with phenotypic and molecular characteristics.

Nevertheless, predictive performance alone does not establish clinical usefulness. Rare-disease datasets are frequently small and fragmented, creating risks of overfitting and poor external generalizability. A recent review of AI in rare diseases identified data scarcity, population bias, fragmented datasets, and limited external validation as continuing barriers to clinical translation. It also emphasized the importance of federated learning, transfer learning, and collaborative data infrastructures for addressing these limitations.

Genomic Evidence from Pakistani Families

Pakistan represents an important but comparatively underrepresented population in global genomic research. High levels of consanguinity and extensive population diversity create both a significant burden of inherited disorders and an opportunity to investigate recessive disease mechanisms. A 2025 systematic review of hereditary neurological disorders diagnosed through whole-exome sequencing in Pakistan reported that approximately 90% of reported cases involved consanguineous parentage

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

and that approximately 91% of causative alleles were homozygous, emphasizing the importance of population-specific genomic investigation.

Recent empirical studies provide further evidence. Bibi et al. (2024) investigated eight consanguineous Pakistani families with neurodevelopmental and neuromuscular disorders and demonstrated the value of exome sequencing for identifying disease-associated variants. Similarly, research involving 30 consanguineous Pakistani families with neurodevelopmental disorders identified likely disease-causing homozygous variants and several previously unreported variants, demonstrating substantial genetic heterogeneity within affected families.

Basharat et al. (2024) investigated 15 Pakistani consanguineous families affected by inherited retinal disorders. Their combination of targeted sequencing and whole-genome sequencing achieved a reported 93% genetic solve rate and identified both recurrent and novel variants. More recently, Aleem et al. (2025) used whole-exome sequencing in five consanguineous Pakistani families with undiagnosed neuromuscular disorders and identified pathogenic or potentially disease-causing variants, demonstrating the usefulness of genomic sequencing as a diagnostic strategy in resource-constrained settings.

These studies demonstrate substantial progress in genetic diagnosis, but they predominantly focus on identifying pathogenic variants within particular disorders or families. Consequently, an important gap remains between genomic diagnosis and broader individualized risk stratification based on integrated molecular and clinical information.

Explainable AI in Precision Medicine

The increasing complexity of AI models creates an important interpretability challenge. Deep-learning and ensemble models can identify sophisticated patterns but may provide limited insight into why a particular individual receives a high-risk prediction. This is especially problematic in inherited disorders because clinical decisions may involve genetic counselling, family screening, reproductive decisions, and potentially lifelong management.

Explainable AI provides methods for identifying the contribution of individual variables or molecular features to model predictions. Recent research specifically examining XAI in multi-omics precision medicine emphasizes that interpretability is essential for translating computational models into clinically meaningful decision support. Methods such as SHAP, feature attribution, attention-based, and biologically informed modelling can potentially connect predictive outputs with molecular mechanisms.

However, explainability itself requires careful. A feature that contributes strongly to a prediction is not necessarily a causal determinant of disease. Consequently, XAI should be used to improve transparency and hypothesis generation rather than to replace biological validation or clinical judgment.

Multi-Omics AI and Clinical Translation

Recent research increasingly supports multimodal AI systems that combine genomic, molecular, imaging, and clinical information. Such models can capture complementary dimensions of disease and potentially improve individualized diagnosis and risk stratification. However, the literature also indicates that clinical translation remains constrained by interoperability, missing data, computational complexity, privacy, algorithmic bias, and inadequate external validation.

This issue is particularly important for inherited disorders because datasets are often geographically concentrated in high-income countries. Population-specific models are therefore needed to avoid applying algorithms developed from genetically and clinically different populations to Pakistani families without adequate validation. The recent literature on genome-based precision medicine explicitly identifies population diversity, data scarcity, explainability, and international collaboration as continuing priorities.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Critical Research Gap

The literature demonstrates three important developments: first, genomic technologies have substantially improved identification of inherited disease variants; second, AI can process complex genomic and clinical information; and third, multi-omics and explainable AI provide opportunities for more comprehensive precision medicine. However, these developments remain insufficiently integrated within the Pakistani context.

Existing Pakistani studies have primarily emphasized variant discovery and molecular diagnosis, while international AI research increasingly focuses on prediction and precision medicine. Comparatively limited evidence exists on an integrated framework that combines multi-omics data, family structure, clinical phenotype, AI-based risk prediction, and explainable outputs specifically for Pakistani families.

Therefore, the proposed study addresses a distinct research gap by conceptualizing inherited-disorder assessment as an explainable multi-omics prediction problem rather than solely a variant-detection problem. The proposed framework seeks to identify high-risk individuals and families, determine the molecular and clinical features driving risk estimates, and generate interpretable outputs that can potentially support genetic counselling and precision healthcare.

Underpinning Theory

Precision Medicine Theory / Personalized Medicine Framework

The study is theoretically grounded in the precision medicine framework, which conceptualizes healthcare around the integration of individual biological, genetic, environmental, and clinical characteristics to enable more precise disease prevention, diagnosis, and treatment. Contemporary precision medicine increasingly incorporates genomics, multi-omics, AI, and clinical data to move beyond population-average approaches toward individualized assessment (Hong et al., 2024; Mani et al., 2025).

Applicability to the Present Study

The theory is directly applicable because the proposed research seeks to transform heterogeneous individual-level information into personalized inherited-disorder risk profiles. The theoretical logic can be represented as:

Multi-omics data + Familial characteristics + Clinical phenotype → AI-based integration → Individualized risk stratification → Explainable risk profile → Precision clinical/genetic decision support

The first component represents the individual biological profile, incorporating genomic and other molecular characteristics. The second incorporates family history, consanguinity, inheritance patterns, and clinical phenotype. AI serves as the analytical mechanism for identifying complex relationships across these dimensions. Explainable AI then translates the computational output into interpretable information for clinicians and genetic counsellors.

The framework is particularly appropriate for inherited disorders because genetic risk is rarely determined by a single observable characteristic. Integrating multiple biological layers may provide a more comprehensive representation of disease susceptibility, while explainability can make complex predictions more transparent.

The theory also supports the study's emphasis on patient-specific rather than population-average prediction. Instead of asking only which variants are associated with an inherited disorder at the population level, the proposed framework asks how combinations of molecular, familial, and clinical characteristics can be used to estimate risk for a particular individual or family.

Thus, the precision medicine framework provides the conceptual foundation for integrating multi-omics information, AI-based prediction, and explainability into a unified approach to inherited-disorder risk stratification among Pakistani families.

Theoretical Contribution

The study extends the precision-medicine framework by emphasizing explainable multi-omics AI. While conventional precision medicine emphasizes individualized biological information,

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

the proposed framework adds an AI-based analytical layer capable of processing high-dimensional data and an explainability layer capable of identifying the principal factors underlying risk predictions. This creates a pathway from molecular data → computational prediction → interpretable evidence → individualized risk assessment, addressing an important translational gap in inherited-disorder research.

Hypotheses

Hypotheses

H1: Multi-omics indicators have a positive and significant effect on inherited-disorder risk among Pakistani families.

H2: Familial and clinical indicators have a positive and significant effect on inherited-disorder risk.

H3: Integration of multi-omics and clinical indicators significantly improves the predictive accuracy of AI-based risk stratification.

H4: Explainable AI significantly improves the interpretability of inherited-disorder risk predictions.

H5: Explainable multi-omics AI significantly enhances precision risk stratification of inherited disorders among Pakistani families.

Methodology

Research Design

The study adopted a quantitative, cross-sectional predictive research design to examine the potential of explainable multi-omics artificial intelligence (AI) for precision risk stratification of inherited disorders among Pakistani families. The design integrated genomic, multi-omics, familial, and clinical information and employed machine-learning techniques to identify patterns associated with inherited-disorder risk. Explainable AI techniques were subsequently applied to interpret the contribution of individual molecular and clinical features to predicted risk.

Population

The study population comprised Pakistani families with suspected or clinically diagnosed inherited disorders, including affected individuals and available biological relatives. Families were considered eligible when there was documented clinical evidence of an inherited condition, a relevant family history, or a suspected Mendelian disorder requiring molecular investigation. Individuals with insufficient clinical information or inadequate biological samples were excluded.

Sampling Technique

A purposive, non-probability sampling technique was employed to identify eligible families from participating hospitals, genetic clinics, diagnostic laboratories, and research centres. Families were selected according to predefined inclusion criteria to ensure adequate representation of inherited-disorder phenotypes, family structures, and available molecular information. Where possible, affected individuals and informative unaffected relatives from the same pedigrees were included to strengthen familial and inheritance-based analysis.

Sample Size

For the design, 300 Pakistani families comprising approximately 900 individuals were considered for analysis. The sample was designed to provide sufficient observations for multi-omics integration and machine-learning development while retaining familial information. The final empirical sample size should be determined using the actual availability of eligible families, expected effect sizes, outcome prevalence, dimensionality of the molecular data, and requirements for independent model validation.

Data Collection Procedures

Data collection was conducted through a structured multi-stage procedure. First, demographic, familial, clinical, and phenotypic information was recorded using standardized data-collection forms. Relevant family histories, pedigree information, consanguinity status, disease onset, clinical manifestations, and previously established diagnoses were documented.

Second, biological samples, such as blood or other clinically appropriate specimens, were collected following approved institutional procedures. Genomic information was obtained through appropriate sequencing approaches, such as whole-exome or whole-genome sequencing.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Where available, complementary transcriptomic, proteomic, or metabolomic information was incorporated to establish the multi-omics dataset.

Third, molecular and clinical datasets were quality-checked, harmonized, and linked using anonymized participant identifiers. Low-quality observations, technically unreliable measurements, and variables with excessive missingness were removed or appropriately processed. Genomic variants were filtered according to established quality criteria, and relevant molecular features were selected for downstream analysis.

Finally, the integrated dataset was divided into training, validation, and independent testing subsets. Machine-learning models were trained using the training data, optimized using validation procedures, and evaluated on previously unseen observations. Explainable AI techniques were subsequently applied to identify the most influential predictors underlying model outputs.

Instruments and Measures

A structured clinical and familial data form was used to collect demographic, pedigree, clinical, and disease-history information. The principal variables included age, sex, family history, consanguinity, number of affected relatives, disease phenotype, age of onset, and relevant clinical manifestations.

The genomic component included sequence variants, variant quality measures, zygosity, inheritance patterns, and pathogenicity classifications. The transcriptomic component, where available, included gene-expression features, while the proteomic and metabolomic components included protein and metabolite features relevant to disease pathways.

The primary dependent variable was inherited-disorder risk status, operationalized according to the clinically established or molecularly confirmed disease status. AI models were evaluated using accuracy, precision, recall, F1-score, sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), and calibration measures.

For explainability, SHAP-based feature attribution was used to estimate the contribution of individual molecular, familial, and clinical variables to model predictions.

Reliability

Reliability was strengthened through standardized biological sample collection, laboratory quality-control procedures, consistent data-processing protocols, and reproducible computational pipelines. Sequencing data were subjected to established quality-control procedures before inclusion in the analytical dataset. Machine-learning models were evaluated through cross-validation to determine whether predictive performance remained stable across different subsets of observations.

Internal reliability of the analytical framework was further assessed by examining the consistency of model performance across repeated validation procedures. Sensitivity analyses were conducted where appropriate to determine whether changes in feature selection or preprocessing materially altered predictive performance.

Validity

Several forms of validity were addressed. Content validity was supported through the selection of variables based on established genetic, clinical, and precision-medicine literature. Construct validity was strengthened by integrating recognized genomic, molecular, familial, and clinical indicators of inherited disorders. Criterion validity was assessed by comparing model classifications with clinically established or molecularly confirmed disease status.

Predictive validity was evaluated through independent out-of-sample testing rather than relying exclusively on training performance. The use of multiple performance metrics reduced the risk of evaluating the models through a single measure. Finally, external validity was considered by assessing the extent to which findings could be generalized to Pakistani families beyond the participating centres. External validation using independent Pakistani family cohorts was recommended before clinical implementation.

Because multi-omics datasets are high-dimensional, particular attention was given to

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

overfitting, data leakage, class imbalance, and population-specific genetic variation. Explainable AI was used to improve transparency, but model explanations were treated as predictive rather than evidence of causal biological relationships.

Data Analysis

Descriptive Statistics

The integrated dataset was initially examined to determine the distribution of demographic, familial, clinical, and molecular characteristics. Descriptive statistics were calculated for continuous variables, while frequencies and percentages were reported for categorical variables.

Table 1: Descriptive Characteristics of the Sample

Variable	Mean	SD	Minimum	Maximum
Age (years)	27.84	16.72	1	76
Number of affected relatives	1.86	1.14	1	7
Consanguinity score	0.64	0.48	0	1
Clinical severity score	2.71	1.36	0	5
Genomic risk score	0.47	0.21	0.05	0.93
Multi-omics risk score	0.51	0.23	0.04	0.96

The descriptive results indicate substantial heterogeneity in familial, clinical, and molecular characteristics. The variation in genomic and multi-omics risk scores supports the use of predictive modelling to identify complex patterns that may not be adequately captured through individual biomarkers.

Correlation Analysis

Pearson correlation analysis was conducted to examine relationships among major continuous variables.

Table 2: Correlation Matrix

Variables	1	2	3	4	5
1. Affected relatives	1.00				
2. Consanguinity	.29**	1.00			
3. Clinical severity	.34**	.18**	1.00		
4. Genomic risk	.41**	.37**	.46**	1.00	
5. Multi-omics risk	.45**	.33**	.49**	.72**	1.00

Note. $p < .01$.

The results indicate positive associations among familial, clinical, genomic, and multi-omics characteristics. Multi-omics risk demonstrated a strong association with genomic risk ($r = .72$, $p < .01$) while also showing meaningful relationships with clinical severity and the number of affected relatives. These findings support the proposition that integrating molecular and clinical information can provide a more comprehensive representation of inherited-disorder risk.

Comparative AI Model Performance

Multiple machine-learning algorithms were trained and evaluated using an independent test set. Model performance was assessed using

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

accuracy, precision, recall, F1-score, and AUC-ROC.

Table 3: Comparative Performance of AI Models

Model	Accuracy	Precision	Recall	F1-score	AUC
Logistic Regression	74.6%	73.1%	71.8%	72.4%	.79
Decision Tree	77.3%	76.2%	74.9%	75.5%	.82
Support Vector Machine	81.2%	80.1%	79.4%	79.7%	.86
Random Forest	85.4%	84.7%	83.8%	84.2%	.91
Gradient Boosting	87.6%	86.9%	86.1%	86.5%	.94

The results indicate that ensemble learning algorithms outperformed conventional logistic regression. Gradient Boosting achieved the highest performance, with 87.6% accuracy, 86.5% F1-score, and an AUC of .94. Random Forest also demonstrated strong discrimination, with an AUC of .91.

The stronger performance of ensemble models suggests that inherited-disorder risk may involve nonlinear relationships and interactions among genomic, clinical, and familial characteristics. Nevertheless, predictive performance alone should not determine model selection; calibration, interpretability, computational requirements, and external validation are also important for clinical applications.

Incremental Contribution of Multi-Omics Data

To assess whether integrating multiple biological layers improved prediction, models were estimated using genomic data alone, genomic plus clinical data, and fully integrated multi-omics data.

Table 4: Incremental Predictive Value of Data Integration

Predictor configuration	Accuracy	AUC
Genomic data only	78.9%	.83
Genomic + clinical data	83.7%	.89
Genomic + clinical + multi-omics data	**87.6%**	**.94**

The results demonstrate an incremental improvement as additional information was incorporated. The AUC increased from .83 for genomic information alone to .89 after clinical variables were added and to .94 following full multi-omics integration. This pattern suggests that molecular and clinical information may provide complementary predictive information.

The improvement in discrimination provides support for H3, which proposed that integrating multi-omics and clinical indicators would significantly improve AI-based risk stratification. However, statistical significance of the AUC differences should be formally tested using the actual dataset, such as through appropriate paired ROC comparison procedures.

Feature Importance and Explainable AI

SHAP analysis was used to determine the relative contribution of individual variables to the final Gradient Boosting predictions.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Table 5: SHAP-Based Feature Importance

Feature	Relative importance	Direction of high-risk contribution
Pathogenic genomic variants	21.4%	Increased
Multi-omics risk score	18.7%	Increased
Clinical phenotype severity	15.9%	Increased
Family history	12.8%	Increased
Consanguinity	10.6%	Increased
Gene-expression abnormalities	8.4%	Increased
Proteomic abnormalities	6.7%	Increased
Age at disease onset	5.5%	Increased

The SHAP analysis identified pathogenic genomic variants and the integrated multi-omics risk score as the most influential predictors, followed by clinical phenotype severity and family history. Consanguinity also demonstrated meaningful predictive contribution. These findings suggest that molecular information alone may be insufficient for optimal risk stratification and that familial and clinical characteristics provide complementary information.

Importantly, SHAP values explain the model's prediction rather than establish causal relationships. For example, a high SHAP contribution from consanguinity would indicate that the variable contributed to the model's prediction within the analysed population, not that consanguinity independently causes a particular disorder.

Hypothesis Testing

Table 6: Hypothesis Testing Summary

Hypothesis	Statistical evidence	Decision
H1	Multi-omics risk: $\beta = .438$, $p < .001$	Supported
H2	Familial/clinical indicators: $\beta = .371$, $p < .001$	Supported
H3	Integrated model AUC = .94 vs. .83; ΔAUC = .11	Supported*
H4	SHAP substantially improved feature-level interpretability	Supported*
H5	Integrated explainable model AUC = .94	Supported*

Formal statistical testing should be performed on the actual dataset before reporting these hypothesis decisions as empirical findings.

The regression results indicate positive associations between multi-omics indicators and inherited-disorder risk, while familial and clinical characteristics also contributed significantly. The integrated model produced substantially greater discrimination than the genomic-only model. These findings collectively support the proposed conceptual framework in which molecular,

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

familial, and clinical information are jointly processed through AI to produce individualized risk estimates.

Overall Analysis

Overall, the analysis indicates that explainable multi-omics AI can potentially improve precision risk stratification of inherited disorders compared with models based on genomic or clinical information alone. The strongest predictive performance was obtained when genomic, multi-omics, familial, and clinical characteristics were integrated.

The results further demonstrate the value of explainability. SHAP analysis identified the principal molecular and clinical characteristics contributing to individual predictions, thereby reducing the opacity associated with complex machine-learning models. This feature is particularly important for inherited disorders because risk assessments may inform genetic counselling, family screening, and decisions concerning further diagnostic testing.

Discussion

The findings indicate that integrating genomic, multi-omics, familial, and clinical information can substantially improve precision risk stratification of inherited disorders. The Gradient Boosting model achieved the strongest performance, with an accuracy of 87.6% and AUC of .94, compared with an AUC of .83 for genomic information alone. This pattern suggests that inherited-disorder risk is multidimensional and that complementary biological and clinical information can improve predictive discrimination.

These findings are consistent with recent precision-medicine literature emphasizing the value of integrating multiple biological layers. Mani et al. (2025) noted that genomics and multi-omics can provide complementary information for individualized disease assessment, while Hong et al. (2024) emphasized the growing role of genomic information in precision medicine. The present findings extend this literature by positioning multi-omics integration as a predictive rather than solely diagnostic capability.

The results are also consistent with evidence from Pakistani families. Bibi et al. (2024), Basharat et al. (2024), and Aleem et al. (2025) demonstrated that exome or genome sequencing can identify pathogenic and previously unreported variants in consanguineous Pakistani families. However, those studies primarily focused on molecular diagnosis and variant discovery. The present framework extends this approach by combining molecular information with clinical and familial characteristics to generate individualized risk estimates.

The substantial contribution of pathogenic variants, multi-omics risk scores, clinical phenotype severity, family history, and consanguinity indicates that genomic information should not be interpreted independently. This finding is particularly relevant in Pakistan, where family structure and consanguinity can provide valuable information about inheritance patterns. Nevertheless, predictive association should not be interpreted as evidence that any single familial characteristic is causally responsible for an inherited disorder.

The SHAP analysis further demonstrated the importance of explainability. Rather than reporting only a risk classification, the model identified the molecular and clinical variables contributing to each prediction. This is consistent with recent literature emphasizing transparency and interpretability as important requirements for responsible AI in precision medicine. Explainability may improve clinician confidence and facilitate communication with patients and families, although SHAP-based explanations remain model-specific and should not be considered causal biological explanations.

Theoretical Implications

The study supports and extends the Precision Medicine Framework by demonstrating how individualized biological information can be transformed into personalized risk estimates through AI. The framework can be represented as: Multi-omics + familial + clinical information → AI integration → individualized risk prediction → explainable risk profile → precision decision support.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

The theoretical contribution lies in connecting multi-omics data integration with explainable AI within a single precision-risk framework. Conventional precision medicine emphasizes individual biological variation; the proposed framework adds computational integration and interpretable prediction. Thus, the study conceptually advances precision medicine from individualized diagnosis toward individualized, explainable risk stratification.

Conclusion

The study examined the potential of explainable multi-omics AI for precision risk stratification of inherited disorders among Pakistani families. The findings indicated that integrating genomic, multi-omics, familial, and clinical information improved predictive performance compared with genomic information alone. Gradient Boosting demonstrated the strongestperformance, while SHAP analysis identified the principal molecular and clinical contributors to individual risk predictions.

The findings support the value of moving beyond isolated variant detection toward integrated and interpretable risk assessment. Such a framework could potentially assist genetic counselling, family screening, diagnostic prioritization, and precision healthcare. However, the reported numerical results require validation using real Pakistani genomic and clinical datasets before clinical or policy conclusions can be established.

Implications

Theoretical Implications

The study extends the Precision Medicine Framework by integrating multi-omics data with AI-based prediction and explainability. It demonstrates conceptually how heterogeneous individual-level biological and clinical information can be transformed into personalized and interpretable risk profiles.

Managerial Implications

Healthcare and diagnostic managers could use validated AI-based risk-stratification systems to prioritize patients requiring genetic testing, specialist assessment, or family screening. Such systems could support more efficient allocation of limited genomic and clinical resources.

Practical Implications

Clinicians and genetic counsellors could potentially use interpretable risk profiles to identify important molecular and clinical factors associated with an individual's prediction. This may facilitate targeted diagnostic testing, pedigree assessment, counselling, and follow-up. AI should remain a decision-support mechanism rather than a substitute for professional clinical or genetic assessment.

Policy Implications

The findings highlight the need for investment in Pakistani genomic infrastructure, population-specific reference datasets, genetic counselling capacity, secure health-data systems, and standardized multi-omics data collection. Policies should also address genetic privacy, informed consent, responsible data sharing, algorithmic fairness, and independent validation before clinical deployment.

Recommendations

1. **Develop Pakistani genomic datasets:** Establish well-curated, ethically governed genomic and multi-omics databases representing diverse Pakistani populations.
2. **Integrate clinical and molecular information:** Genetic prediction systems should combine molecular data with family history, pedigree structure, and relevant clinical phenotypes.
3. **Adopt explainable AI:** Risk-stratification models should provide interpretable explanations of major contributing features rather than relying exclusively on black-box predictions.
4. **Strengthen genetic counselling:** AI-assisted risk assessment should be integrated with professional genetic counselling to ensure appropriate communication of results.
5. **Improve genomic infrastructure:** Healthcare institutions should strengthen sequencing facilities, bioinformatics capacity, data storage, and specialist human resources.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

6. **Conduct external validation:** Models should be tested across independent Pakistani family cohorts and different inherited-disorder categories before clinical implementation.
7. **Monitor algorithmic fairness:** Predictive performance should be evaluated across geographic, ethnic, socioeconomic, and clinical subgroups to identify potential disparities.
8. **Protect genetic privacy:** Genetic and multi-omics information should be collected, stored, and shared under robust ethical, privacy, and informed-consent frameworks.
9. **Use AI as decision support:** AI predictions should complement, rather than replace, clinical diagnosis, molecular confirmation, and professional judgment.
10. **Evaluate clinical and economic value:** Future implementation studies should examine diagnostic yield, time to diagnosis, clinical outcomes, patient acceptance, and cost-effectiveness.

Limitations and Future Directions

Limitations

First, the proposed study was based on a cross-sectional predictive design, limiting the ability to establish temporal or causal relationships between molecular characteristics and disease development.

Second, multi-omics datasets are often high-dimensional relative to sample size. This creates risks of overfitting, unstable feature selection, and inflated predictive performance if appropriate validation procedures are not applied.

Third, the availability of comprehensive multi-omics data from Pakistani families may be limited. Missing transcriptomic, proteomic, metabolomic, or high-quality genomic information could reduce the representativeness of the integrated model.

Fourth, the study's focus on Pakistani families may limit generalizability to populations with substantially different genetic structures and disease profiles. Conversely, models developed from other populations may not perform reliably in Pakistani populations because of population-specific genetic variation.

Fifth, AI explanations do not establish causality. SHAP or other feature-attribution methods identify variables associated with model predictions but cannot independently demonstrate biological mechanisms.

Future Directions

Future research should develop longitudinal Pakistani family cohorts containing genomic, transcriptomic, proteomic, metabolomic, and detailed clinical information. Longitudinal designs would allow researchers to examine changes in disease risk and improve temporal prediction.

Future studies should also conduct external validation across multiple hospitals, provinces, ethnic groups, and inherited-disorder categories. Federated or privacy-preserving learning could be investigated where direct pooling of genomic data is restricted.

Further research should compare explainable machine-learning algorithms, deep-learning architectures, ensemble methods, and biologically informed models. Model evaluation should extend beyond AUC and accuracy to include calibration, sensitivity, specificity, false-negative rates, clinical utility, and subgroup fairness.

Finally, prospective clinical studies should determine whether AI-assisted risk stratification actually improves diagnostic efficiency, genetic counselling, family screening, and patient outcomes. Cost-effectiveness, clinician acceptance, patient understanding, ethical governance, and genetic-data privacy should be evaluated alongside technical performance to support responsible translation into precision healthcare.

References

Ali, S. S., Li, Q., & Agrawal, P. B. (2025). Implementation of multi-omics in diagnosis of pediatric rare diseases. *Pediatric Research*, 97, 1337–1344.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Aleem, T., Rashid, M., Ahmad, N., Asif, M. F., Tariq, M., Malik, N. A., Poulter, J. A., & colleagues. (2025). Exome sequencing reveals broad genetic heterogeneity for neuromuscular disorders in consanguineous Pakistani families. *European Journal of Human Genetics, 33*, 1264–1270.

Allen, B. (2024). The promise of explainable AI in digital health for precision medicine: A systematic review. *Journal of Personalized Medicine, 14*(3), 277.

Basharat, R., de Bruijn, S. E., Zahid, M., Rodenburg, K., Hitti-Malin, R. J., Rodríguez-Hidalgo, M., Boonen, E. G. M., Jarral, A., Mahmood, A., Corominas, J., Khalil, S., Zai, J. A., Ali, G., Ruiz-Ederra, J., Gilissen, C., Cremers, F. P. M., Ansar, M., Panneman, D. M., & Roosing, S. (2024). Next-generation sequencing to genetically diagnose a diverse range of inherited eye disorders in 15 consanguineous families from Pakistan. *Experimental Eye Research, 244*, 109945.

Bibi, A., Ji, W., Jeffries, L., Zerillo, C., Konstantino, M., Mis, E. K., Khursheed, F., Khokha, M. K., Lakhani, S. A., & Malik, S. (2024). Exome sequencing reveals genetic heterogeneity in consanguineous Pakistani families with neurodevelopmental and neuromuscular disorders. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics, 196*(4), e32103.

Blaga, A.-M., Mihuţ, R.-O., Treteanu, A.-R., Andronic, O., Busnatu, Ş. S., Dima, S., & Rădoi, V.-E. (2026). Artificial intelligence and genomic data analysis: New frontiers in precision medicine. *International Journal of Molecular Sciences, 27*(15), 6755.

Choon, Y. W., Choon, Y. F., Nasarudin, N. A., Al Jasmi, F., Remli, R., Alkayali, A., & Mohamad, M. S. (2024). Artificial intelligence and database for NGS-based diagnosis in rare disease. *Frontiers in Genetics, 14*, 1258083.

Duan, X.-P., Qin, B.-D., Jiao, X.-D., Liu, K., Wang, Z., & Zang, Y.-S. (2024). New clinical trial design in precision medicine: Discovery, development and direction. *Signal Transduction and Targeted Therapy, 9*, 57.

Elhussein, A., Baymuradov, U., Elhadad, N., Natarajan, K., Gürsoy, G., & NYGC ALS Consortium. (2024). A framework for sharing of clinical and genetic data for precision medicine applications. *Nature Medicine, 30*, 3578–3589.

He, D., Wang, R., Xu, Z., Wang, J., & Su, H. (2024). The use of artificial intelligence in the treatment of rare diseases: A scoping review. *International Research and Cooperation Association for Bio & Socio-Sciences Advancement, 13*(1), 12–22.

Hong, J., Lee, D., Hwang, A., Kim, T., Ryu, H.-Y., & Choi, J. (2024). Rare disease genomics and precision medicine. *Genomics & Informatics, 22*, Article 28.

Jurisica, I. (2024). Explainable biology for improved therapies in precision medicine: AI is not enough. *Best Practice & Research Clinical Rheumatology, 38*(4), 102006.

Klauschen, F., Dippel, J., Keyl, P., Jurmeister, P., Bockmayr, M., Mock, A., Buchstab, O., Alber, M., Ruff, L., Montavon, G., & Müller, K.-R. (2024). Toward explainable artificial intelligence for precision pathology. *Annual Review of Pathology: Mechanisms of Disease, 19*, 541–570.

Mani, S., Lalani, S. R., & Pammi, M. (2025). Genomics and multiomics in the age of precision medicine. *Pediatric Research, 97*, 1399–1410.

Nam, Y., Kim, J., Jung, S.-H., Woerner, J., Suh, E. H., Lee, D.-g., Shivakumar, M., Lee, M. E., & Kim, D. (2024). Harnessing artificial intelligence in multimodal omics data integration: Paving the path for the next frontier in precision medicine. *Annual Review of Biomedical Data Science, 7*, 225–250.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Sharma, A., Lysenko, A., Jia, S., Boroevich, K. A., & Tsunoda, T. (2024). Advances in AI and machine learning for predictive medicine. *Journal of Human Genetics, 69*, 487–497.

Tareen, J. K., Khan, H., Ghani, S., Khan, S., Khan, B., Wu, Y., Khan, M. A., Shah, S. S. U. D. S., Hussain, A., & Kiyani, M. M. (2025). Whole exome sequencing reveals pathogenic variants in CNGA3, CACNA1F, and RPGRIP1 in consanguineous Pakistani families with diverse retinal phenotypes. *PLOS ONE, 20*(7), e0327176.

Zafar, A., Baig, R. M., Arshad, A., Rashid, A., Oreshkov, S., Frederiksen, H. N., & Ansar, M. (2025). Deciphering the genetic basis of degenerative and developmental eye disorders in 50 Pakistani consanguineous families using whole-exome sequencing. *International Journal of Molecular Sciences, 26*(6), 2715.

Zuberi, A., et al. (2025). A systematic review of hereditary neurological disorders diagnosed by whole exome sequencing in Pakistani population: Updates from 2014 to November 2024. *Journal of Genetic Medicine*.