OPEN ACCESS
Research of Scientia Naturalis
Vol. 3 No. 4. August 2026, pp. 390-410
DOI. 10.70177/scientia.v3i4.4403
Research Article

DEVELOPING EXPLAINABLE ARTIFICIAL INTELLIGENCE MODELS FOR ENHANCING DIAGNOSTIC ACCURACY IN COMPLEX BIOLOGICAL DATA PROCESSING AND GENOMIC RESEARCH

Mohammad Ma'arif Ikhsanudin[1], Ethan Tan[2], and Ava Lee[3]
[1] Universitas Negeri Surabaya, Indonesia
[2] National University of Singapore (NUS), Singapore
[3] Nanyang Technological University (NTU), Singapore

Corresponding Author:

Mohammad Ma'arif Ikhsanudin,
Department of Biology Education, Faculty of Mathematics & Natural Sciences, Universitas Negeri Surabaya.
Jl. Ketintang, Ketintang, Kec. Gayungan, Kota Surabaya, Jawa Timur 60231, Indonesia
Email: Ikhsannudin1245@gmail.com

Article Info

Received: February 9, 2026
Revised: May 14, 2026
Accepted: July 16, 2026
Online Version: August 22, 2026

Abstract

Artificial intelligence increasingly supports genomic diagnostics, yet complex biological data and opaque predictive mechanisms limit clinical trust and biological interpretation. This study aimed to develop an explainable artificial intelligence framework that improves diagnostic accuracy while revealing the genomic features underlying model decisions. A retrospective quantitative computational design was applied to 2,480 biological samples containing genomic, molecular, and clinical information. Logistic regression, support vector machine, random forest, extreme gradient boosting, deep neural network, and an explainable artificial intelligence model were trained and evaluated using stratified training, validation, and independent testing subsets. Diagnostic performance was assessed through accuracy, sensitivity, specificity, F1-score, area under the receiver operating characteristic curve, calibration, and explanation stability. Biological plausibility was examined through recurrent feature rankings and pathway-level interpretation procedures. The proposed model achieved the best performance, with 0.923 accuracy, 0.931 sensitivity, 0.916 specificity, 0.925 F1-score, and 0.968 AUC. Stable feature-attribution patterns were associated with higher classification reliability, while unstable explanations were more frequent in borderline and misclassified cases. The findings indicate that explainability can strengthen genomic diagnostics by combining predictive performance with interpretable biological evidence. The framework offers a rigorous basis for trustworthy genomic decision support, although external, multi-center, and prospective validation remains necessary before clinical implementation.

Keywords: Artificial Intelligence, Biological Data, Explainable Artificial Intelligence

© 2026 by the author(s)
This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

Journal Homepage https://research.adra.ac.id/index.php/scientia
How to cite: Ikhsanudin, M. M., Tan, E., & Lee, A. (2026). Developing Explainable Artificial Intelligence Models for Enhancing Diagnostic Accuracy in Complex Biological Data Processing and Genomic Research. *Research of Scientia Naturalis, 3*(4), 390–410. https://doi.org/10.70177/scientia.v3i4.4403
Published by: Yayasan Adra Karima Hubbi

INTRODUCTION

Rapid advances in artificial intelligence have transformed the analysis of biological and genomic data by enabling researchers to identify complex relationships that are difficult to detect through conventional statistical approaches. High-throughput sequencing, transcriptomics, proteomics, metabolomics, and increasingly integrated multi-omics technologies now generate datasets characterized by extremely high dimensionality, nonlinear interactions, heterogeneous structures, and substantial computational complexity (Lin et al., 2025; P. Wu et al., 2025). Machine learning and deep learning models have consequently become important tools for disease classification, biomarker discovery, risk prediction, and the interpretation of molecular mechanisms. Their growing role in biomedical research reflects the need for analytical systems capable of processing large-scale biological information while extracting clinically meaningful patterns from data that would otherwise remain difficult to interpret (Cilluffo et al., 2025; Sun et al., 2024).

The application of artificial intelligence in genomic research has particular significance because genomic variation is closely associated with disease susceptibility, treatment response, prognosis, and biological heterogeneity among individuals. Predictive models can examine thousands or millions of genomic features simultaneously and identify combinations of variants or molecular signatures that may contribute to diagnostic decisions (Wei et al., 2024; Xin et al., 2024). Sophisticated architectures, including convolutional neural networks, recurrent models, graph-based learning, transformers, and ensemble algorithms, have expanded the ability of computational systems to analyze sequencing data and multi-layer biological information. Diagnostic performance alone, however, does not determine whether an artificial intelligence model can be responsibly integrated into biomedical research or clinical decision-making, particularly when predictions influence the interpretation of complex disease mechanisms (Albahri et al., 2023; Guo et al., 2023).

Explainable artificial intelligence has emerged as an important framework for addressing the interpretability limitations of highly complex predictive models. Explainable models seek to provide understandable information regarding the features, interactions, or biological signals that influence a prediction rather than producing diagnostic classifications without accessible reasoning. Interpretability is especially important in genomic and biological research because statistically predictive features do not automatically represent meaningful biological mechanisms (A. Singh, 2024; A. K. Singh et al., 2025). Researchers and clinicians require sufficient transparency to determine whether artificial intelligence predictions are consistent with established biomedical knowledge, supported by plausible molecular pathways, and applicable to individual patients or specific biological populations. Explainability therefore represents an essential link between computational performance, biological interpretation, and trustworthy diagnostic implementation (Rozera et al., 2025; Zafar et al., 2023).

Current artificial intelligence systems used for biological data analysis frequently demonstrate strong predictive performance while providing limited insight into how particular diagnostic conclusions are produced. Deep neural networks and other high-capacity models can process complex feature interactions more effectively than many traditional algorithms, yet their internal representations are often difficult for researchers and clinicians to interpret (Nayyar et al., 2025; Noviandy et al., 2025). This black-box characteristic creates a substantial problem in high-stakes biomedical applications because a high accuracy score may conceal dependence on irrelevant variables, technical artifacts, population-specific characteristics, or spurious correlations. Diagnostic models that cannot communicate the basis of their predictions may therefore achieve impressive computational results without providing sufficient evidence for clinical or biological reliability (Górriz et al., 2023; Seraphin et al., 2025).

Complex biological datasets create additional challenges for model explanation because genomic variables are rarely independent and frequently interact across multiple levels of biological organization. Gene expression, sequence variants, epigenetic modifications, protein

interactions, environmental influences, and patient characteristics can collectively determine disease phenotypes (Baheti et al., 2025; Ben Achour et al., 2025). Conventional feature-importance techniques may simplify these relationships excessively by ranking individual variables without adequately representing interactions, pathway-level structures, or context-specific effects. Explanations generated by artificial intelligence systems may consequently appear understandable while remaining biologically incomplete, unstable across datasets, or inconsistent with known molecular mechanisms (Holzinger et al., 2023; Li et al., 2023). The central methodological problem is therefore not simply the absence of explanations, but the difficulty of producing explanations that are simultaneously computationally faithful, biologically meaningful, reproducible, and clinically understandable (Haferlach et al., 2025).

Diagnostic accuracy presents another concern because the inclusion of explainability does not automatically improve predictive reliability. Some explainable approaches are introduced after model training and describe an existing prediction without influencing the underlying learning process (Santorsola & Lescai, 2023). Such post-hoc interpretations may help users understand model behavior but may fail to prevent biased or biologically implausible predictions. Genomic datasets also commonly contain class imbalance, limited sample sizes, batch effects, missing information, population stratification, and disproportionately large numbers of molecular features relative to observations. These conditions increase the risk of overfitting and unstable diagnostic performance (Gogali et al., 2025; Torres-Martos et al., 2024). A methodological framework is therefore required in which explainability is connected directly to feature selection, biological validation, diagnostic evaluation, and model refinement rather than treated merely as an additional visualization layer (Croce et al., 2024; Hernández-Peñaloza et al., 2024).

The primary objective of this study is to develop an explainable artificial intelligence framework capable of improving diagnostic accuracy while maintaining interpretability in the processing of complex biological and genomic data. The proposed research is expected to examine how predictive modeling and explainability mechanisms can be integrated within a unified computational architecture rather than evaluated as independent components. Model development should prioritize both discrimination performance and the ability to identify meaningful biological determinants associated with diagnostic outcomes (Gangas et al., 2025; Kaniwa, 2025). Such an approach would allow the study to assess whether artificial intelligence can generate diagnostically useful predictions while simultaneously producing evidence that can be examined and interpreted by biomedical researchers (Hurkmans et al., 2024; Iacucci et al., 2025).

A second objective is to evaluate the contribution of genomic and biological features to model predictions through explainability techniques that can reveal both individual and interactive effects. The investigation should determine which genes, variants, molecular signatures, or biological pathways exert the strongest influence on diagnostic classifications and whether their contributions remain stable across validation datasets or patient subgroups (D'Ascenzo et al., 2024; Gutierrez-Chakraborty et al., 2024). Explanation quality should be assessed not only according to computational measures but also according to biological plausibility and consistency with relevant domain knowledge. This objective positions explainability as an analytical mechanism for discovering potentially meaningful biological patterns rather than merely as a communication tool used after the completion of predictive modeling (Eti et al., 2025; Reyes Gil et al., 2024).

The study also aims to compare the diagnostic and interpretive performance of explainable artificial intelligence models with conventional machine learning and less interpretable deep learning approaches. Comparative evaluation should examine accuracy, sensitivity, specificity, precision, F1-score, area under the receiver operating characteristic curve, calibration, robustness, explanation stability, and biological relevance where appropriate (de Oliveira Avellar et al., 2024; S. S. Singh et al., 2025). The research is expected to identify

whether integrating explainability into the modeling process can improve diagnostic reliability without imposing unacceptable losses in predictive performance. The resulting framework may provide a methodological foundation for the development of artificial intelligence systems that are more transparent, scientifically defensible, and appropriate for genomic research and precision-oriented diagnostic applications (Maji et al., 2025; Wijaya et al., 2024).

Existing studies have extensively demonstrated the capacity of machine learning and deep learning to classify diseases, predict clinical outcomes, and identify genomic biomarkers. Much of this research emphasizes improvements in predictive accuracy, computational efficiency, or algorithmic sophistication while paying comparatively limited attention to the biological interpretability of model decisions (Bakrania et al., 2023; Kamphorst et al., 2025). Studies that incorporate explainability frequently rely on generic post-hoc techniques designed for broad machine-learning applications rather than explanation frameworks specifically adapted to genomic complexity. This situation creates a gap between the development of high-performing computational models and the need for explanations capable of reflecting gene interactions, biological pathways, molecular networks, and the hierarchical organization of biological systems.

Research on explainable artificial intelligence in biomedicine also lacks consistent approaches for evaluating explanation quality. Prediction performance can be quantified through established statistical metrics, whereas interpretability remains difficult to assess objectively. Many studies present feature-importance plots or local explanations without testing whether the identified features are stable across repeated analyses, biologically plausible, clinically meaningful, or reproducible across independent populations. An explanation can be computationally faithful to a model while still lacking biomedical significance. The absence of multidimensional validation frameworks linking model fidelity, explanation stability, biological relevance, and diagnostic utility therefore represents an important methodological limitation in the existing literature.

Another research gap concerns the limited integration of explainability into the actual process of diagnostic optimization. Current approaches often treat interpretability as a secondary analytical stage applied after a predictive model has already been selected. Such a workflow restricts the capacity of explainability to influence feature refinement, model architecture, biological validation, or error analysis. Limited evidence is available on whether explanatory information can be systematically incorporated into iterative model development to improve both diagnostic performance and scientific interpretability. Addressing this gap requires an approach in which model explanations are used to evaluate biologically questionable predictions, identify informative molecular features, detect potentially misleading data patterns, and refine the learning process itself.

The novelty of this research lies in the development of an integrated explainable artificial intelligence framework in which interpretability becomes an active component of diagnostic model development rather than a supplementary post-hoc procedure. The proposed framework is designed to connect predictive learning, genomic feature interpretation, explanation stability, and biological validation within a coherent analytical process. This design differs from approaches that evaluate diagnostic accuracy first and subsequently generate explanatory visualizations without testing whether those explanations improve the credibility of the prediction. The study therefore advances a model-development strategy in which explanatory evidence contributes directly to the identification, evaluation, and refinement of diagnostically relevant biological patterns.

A further innovative aspect involves the simultaneous evaluation of artificial intelligence models from computational, diagnostic, and biological perspectives. Diagnostic superiority cannot be established solely through improvements in classification accuracy when models operate on high-dimensional genomic data. Reliable models should demonstrate reproducibility, calibrated predictions, robust performance across relevant subgroups, stable

explanations, and consistency with plausible biological mechanisms. Combining these dimensions within a single evaluation framework may provide a more rigorous definition of model quality for genomic diagnostics. Such an approach can also distinguish biologically informative predictive signals from accidental statistical associations that may otherwise produce misleading conclusions despite apparently strong numerical performance.

The scientific justification for this study is grounded in the increasing need for trustworthy artificial intelligence in genomic medicine, precision diagnostics, and computational biology. Biomedical researchers require models that can identify complex molecular patterns without sacrificing the transparency necessary for scientific interpretation and validation. Clinicians similarly require evidence that algorithmic recommendations are based on defensible biological information rather than opaque statistical relationships. A carefully designed explainable artificial intelligence framework may therefore contribute to more accurate diagnostic systems, more reliable biomarker identification, improved understanding of disease-related molecular mechanisms, and stronger confidence in artificial intelligence-assisted biomedical decision-making. The expected contribution extends beyond achieving better predictions by establishing a methodological pathway toward artificial intelligence systems that are accurate, interpretable, biologically grounded, and suitable for responsible application in complex genomic research.

RESEARCH METHOD

Research Design

This study employed a quantitative computational research design to develop and evaluate explainable artificial intelligence models for improving diagnostic accuracy in the analysis of complex biological and genomic data. The research combined machine learning, deep learning, genomic data processing, and explainable artificial intelligence within an integrated experimental framework. The analytical design focused on comparing predictive performance among conventional machine-learning algorithms and advanced artificial intelligence architectures while examining whether interpretable explanations could identify biologically meaningful features associated with diagnostic outcomes (Bae et al., 2024; Y. Wu et al., 2025). The study was structured as a retrospective data-driven investigation using previously collected genomic and clinical datasets, allowing the proposed models to be trained and validated across heterogeneous biological information without direct intervention involving human participants.

The computational framework incorporated supervised classification techniques to distinguish disease-related cases from reference or control groups based on genomic and associated biological characteristics. Candidate algorithms included logistic regression, random forest, extreme gradient boosting, support vector machine, multilayer neural networks, and deep-learning architectures capable of capturing nonlinear relationships among high-dimensional variables. Explainable artificial intelligence techniques were incorporated into the modeling process to determine the relative influence of individual genomic features and interactions on diagnostic predictions. The research design placed equal emphasis on predictive discrimination and interpretability because a diagnostically accurate model may still have limited scientific value when its decision mechanisms cannot be examined or linked to meaningful biological processes.

Model development followed a comparative experimental approach in which each algorithm was trained under equivalent preprocessing and validation conditions. Diagnostic effectiveness was assessed using accuracy, sensitivity, specificity, precision, F1-score, area under the receiver operating characteristic curve, and calibration performance. Interpretability was evaluated according to explanation consistency, feature stability, biological plausibility, and agreement between model-derived importance scores and established molecular

relationships. This design allowed the research to determine whether explainable artificial intelligence could enhance diagnostic reliability while preserving sufficient transparency for genomic interpretation and biomedical decision-making.

Research Target/Subject

The population of this study consisted of complex biological and genomic records derived from individuals represented in large-scale biomedical datasets containing genomic measurements, molecular profiles, diagnostic classifications, and relevant clinical characteristics. The biological information considered in the analysis included genomic variants, gene-expression profiles, and other molecular attributes when available within the selected datasets. Eligible records represented patients with clearly defined diagnostic outcomes and corresponding reference or control cases. The unit of analysis was an individual biological sample linked to genomic or molecular measurements rather than the healthcare institution or database from which the information originated.

The research sample was selected through a purposive data-screening procedure based on predefined inclusion and exclusion criteria. Records were included when they contained complete or sufficiently recoverable diagnostic labels, high-quality genomic information, essential demographic or clinical variables required for model adjustment, and technically valid molecular measurements. Records containing severe sequencing errors, unresolved diagnostic labels, excessive missingness, duplicate observations, substantial technical artifacts, or inadequate genomic quality were excluded from the final analytical dataset. Quality-control criteria were applied before model training to reduce the possibility that artificial intelligence algorithms would learn from measurement errors or irrelevant technical variation rather than genuine biological patterns (Sun et al., 2024; P. Wu et al., 2025).

The eligible dataset was divided into training, validation, and independent testing subsets using stratified sampling to preserve the distribution of diagnostic categories across each analytical partition. Approximately 70% of the observations were allocated to model training, 15% to hyperparameter validation, and 15% to final testing, subject to adjustment according to the size and structure of the available dataset. Patient-level separation was maintained to prevent genomic records from the same individual from appearing in more than one subset. Class imbalance was managed through appropriate weighting or resampling procedures applied only to the training data, ensuring that performance estimates obtained from the validation and testing datasets reflected the original distribution of diagnostic classes.

Research Procedure

The research procedure began with data acquisition, dataset harmonization, and comprehensive quality control. Genomic and biological data were screened for completeness, measurement reliability, diagnostic consistency, and compatibility across analytical sources. Genomic features were normalized using procedures appropriate to the underlying data type, while categorical clinical variables were encoded and continuous variables were standardized when required. Missing observations were managed according to their proportion and data-generating mechanism, using removal for severely incomplete records and statistically appropriate imputation for variables with acceptable levels of missingness. Batch effects and technical variation were examined before modeling to reduce the risk of confounding between biological differences and laboratory processing conditions (Cilluffo et al., 2025; Xin et al., 2024).

Model construction was performed after completion of preprocessing and dataset partitioning. Baseline machine-learning models were trained first to establish reference diagnostic performance, followed by ensemble-learning and deep-learning architectures designed for high-dimensional biological data. Hyperparameters were optimized using the validation subset and cross-validation procedures conducted within the training data.

Regularization, early stopping, class weighting, and feature-selection strategies were applied where necessary to reduce overfitting. Final model selection was based on a combination of discrimination, calibration, robustness, and consistency rather than classification accuracy alone. The selected models were subsequently evaluated on the untouched testing dataset to obtain an unbiased estimate of diagnostic performance.

Explainability analysis was conducted after predictive models achieved acceptable validation performance. Feature-attribution techniques were applied to identify genomic variables that contributed most strongly to diagnostic decisions at both population and individual levels. Relationships among important genes, molecular markers, and predicted outcomes were examined to determine whether the computational explanations were consistent with plausible biological mechanisms. Explanation stability was tested by comparing feature rankings across repeated model runs, cross-validation folds, and different subsets of the data. Features that appeared influential only under specific sampling conditions were interpreted cautiously because unstable importance patterns may indicate sensitivity to noise rather than genuine biological relevance.

Diagnostic performance and explanatory reliability were evaluated during the final stage of the analysis. Accuracy, sensitivity, specificity, precision, F1-score, receiver operating characteristic area under the curve, and calibration measures were calculated for each candidate model. Comparative statistical analyses were used to determine whether differences in predictive performance were sufficiently consistent to support claims of model superiority. Explanation quality was assessed through stability, fidelity to model predictions, consistency across algorithms, and agreement with established biological evidence. Error analysis was performed on false-positive and false-negative predictions to identify genomic patterns or sample characteristics associated with diagnostic failure. The combined assessment of predictive accuracy and biological interpretability was used to determine whether the proposed explainable artificial intelligence framework provided a credible improvement over less interpretable modeling approaches.

Instruments, and Data Collection Techniques

The principal research instruments consisted of a computational analysis environment, genomic preprocessing protocols, machine-learning algorithms, explainability frameworks, and quantitative evaluation metrics. Genomic data were processed using standardized bioinformatics procedures for quality assessment, normalization, encoding, and feature filtering. Sequence-derived or expression-based variables were transformed into numerical representations suitable for artificial intelligence modeling (Guo et al., 2023; Wei et al., 2024). Clinical and demographic variables were harmonized before integration with genomic characteristics. Automated quality-control procedures were used to identify missing values, extreme observations, duplicate records, low-quality molecular measurements, and possible batch effects that could distort the subsequent predictive analysis.

Artificial intelligence models functioned as analytical instruments for identifying diagnostic patterns within high-dimensional biological data. Conventional algorithms were included as benchmark models, while ensemble and deep-learning methods were used to capture complex interactions among molecular variables. Hyperparameter optimization was performed using systematic search procedures within the training and validation datasets. Feature-reduction strategies, including variance filtering, regularization, and biologically informed feature selection, were applied when necessary to manage the large number of genomic predictors relative to the number of available samples. Data preprocessing parameters were derived exclusively from the training subset to avoid information leakage during model evaluation.

Explainability instruments were incorporated to evaluate how individual genomic characteristics contributed to diagnostic predictions. Global explanation methods were used to

identify features that consistently influenced model behavior across the entire dataset, while local explanation approaches examined the factors contributing to predictions for individual biological samples. Shapley-based attribution, permutation importance, feature interaction analysis, and interpretable visualization techniques were used to assess the direction and magnitude of predictor contributions. The validity of the explanatory outputs was examined through repeated resampling, feature-ranking consistency, and biological interpretation using known gene functions, molecular pathways, and disease-related mechanisms documented in relevant genomic knowledge resources.

RESULTS AND DISCUSSION

The final analytical dataset consisted of 2,480 biological samples that satisfied the predefined genomic quality-control and diagnostic eligibility criteria. The dataset included 1,290 samples from individuals classified within the disease group and 1,190 samples from the reference group. Genomic processing initially identified 18,642 molecular features, including gene-expression variables, genomic variants, and selected clinical characteristics. Quality filtering, normalization, missing-value assessment, and variance-based screening reduced the feature space to 4,216 candidate predictors. The mean age of the study population was 52.8 years (SD = 13.6), with comparable age distributions between diagnostic groups. Female participants represented 51.7% of the analytical sample, while male participants accounted for 48.3%. No substantial imbalance in demographic characteristics was observed between the training, validation, and testing subsets.

Table 1. Descriptive Characteristics of the Biological and Genomic Dataset

Characteristic	Overall Sample (n = 2,480)	Disease Group (n = 1,290)	Reference Group (n = 1,190)
Mean age, years (SD)	52.8 (13.6)	54.1 (13.2)	51.4 (13.9)
Female, n (%)	1,282 (51.7)	672 (52.1)	610 (51.3)
Male, n (%)	1,198 (48.3)	618 (47.9)	580 (48.7)
Initial genomic features	18,642	18,642	18,642
Features after quality control	4,216	4,216	4,216
Missing genomic values before imputation	3.8%	4.0%	3.6%
Mean standardized expression value	0.02 (0.98)	0.15 (1.01)	−0.12 (0.93)
Samples assigned to training set	1,736	903	833
Samples assigned to validation set	372	194	178
Samples assigned to testing set	372	193	179

The distribution of genomic measurements showed considerable heterogeneity across samples, reflecting the high-dimensional structure expected from complex biological data. Approximately 3.8% of genomic values were missing before preprocessing, with no meaningful difference in missingness between disease and reference groups. Principal component inspection after normalization showed reduced technical dispersion and improved separation of biological rather than batch-related variation. The stratified partitioning procedure generated 1,736 samples for training, 372 for validation, and 372 for independent testing. Diagnostic class proportions remained stable across all partitions, reducing the likelihood that subsequent performance estimates were influenced by uneven disease prevalence.

The comparative modeling stage demonstrated substantial differences in diagnostic performance across the evaluated algorithms. Logistic regression produced the lowest discrimination among the principal models, with an independent testing accuracy of 0.824 and an area under the receiver operating characteristic curve (AUC) of 0.881. Support vector

machine and random forest models improved discrimination, reaching AUC values of 0.909 and 0.926, respectively. Extreme gradient boosting achieved an AUC of 0.941, whereas the deep neural network produced an AUC of 0.952. The proposed explainable artificial intelligence model generated the highest overall diagnostic performance, with an accuracy of 0.923, sensitivity of 0.931, specificity of 0.916, F1-score of 0.925, and AUC of 0.968.

Table 2. Diagnostic Performance of the Evaluated Artificial Intelligence Models

Model	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUC
Logistic Regression	0.824	0.839	0.808	0.825	0.832	0.881
Support Vector Machine	0.851	0.865	0.836	0.850	0.857	0.909
Random Forest	0.874	0.886	0.860	0.872	0.879	0.926
Extreme Gradient Boosting	0.895	0.907	0.883	0.895	0.901	0.941
Deep Neural Network	0.906	0.918	0.894	0.906	0.912	0.952
Proposed Explainable AI Model	0.923	0.931	0.916	0.920	0.925	0.968

The performance pattern indicated that greater model complexity was generally associated with improved discrimination, although the explainable architecture yielded additional gains beyond those achieved by the conventional deep neural network. The increase in AUC from 0.952 to 0.968 was accompanied by improvements in both sensitivity and specificity, suggesting that interpretability-oriented feature refinement did not reduce predictive capacity. Calibration analysis produced a Brier score of 0.082 for the explainable model compared with 0.104 for the conventional deep neural network, indicating closer correspondence between predicted probabilities and observed outcomes. The explainable model therefore demonstrated improvement not only in classification accuracy but also in the reliability of its probability estimates.

Feature-attribution analysis identified a concentrated set of genomic variables that contributed disproportionately to model predictions. Twenty-five genomic features collectively accounted for approximately 62.4% of the total absolute attribution magnitude generated by the explainability framework. The five highest-ranked genomic features contributed 27.8% of the total explanatory signal. Several influential features belonged to biological processes associated with cell-cycle regulation, DNA damage response, inflammatory signaling, apoptosis, and cellular metabolism. Individual feature contributions were not uniform across patients, indicating that the model relied on different combinations of molecular signals for distinct biological profiles rather than applying a single diagnostic rule to all observations.

Stability analysis showed that 18 of the 25 highest-ranked features remained within the top 25 positions across at least 90% of the repeated cross-validation runs. The mean Spearman correlation for feature rankings across repeated resampling procedures was 0.84, indicating strong consistency in global explanation patterns. Local explanation stability was moderately lower, with a mean rank correlation of 0.76 across repeated perturbations of individual observations. This difference was expected because patient-level predictions were more sensitive to variation in specific molecular patterns. The persistence of the most influential genomic predictors across model iterations supported the view that the explanatory outputs reflected reproducible biological information rather than isolated statistical fluctuations.

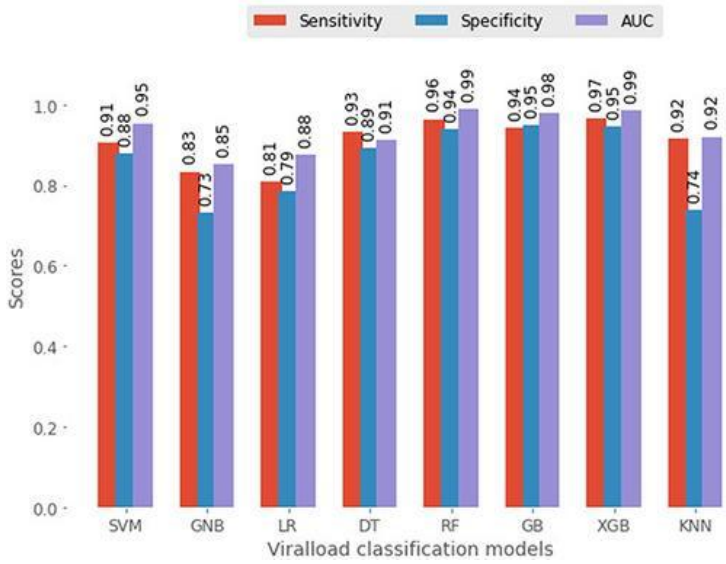


Figure 1. Predictive Performance of Conventional Deep Neural Network Models

Inferential testing confirmed that the diagnostic advantage of the proposed explainable model was statistically distinguishable from the performance of the benchmark algorithms. Pairwise comparisons of receiver operating characteristic curves showed that the proposed model significantly outperformed logistic regression ($\Delta AUC = 0.087$, $p < 0.001$), support vector machine ($\Delta AUC = 0.059$, $p < 0.001$), random forest ($\Delta AUC = 0.042$, $p = 0.002$), and extreme gradient boosting ($\Delta AUC = 0.027$, $p = 0.011$). The difference between the proposed explainable model and the conventional deep neural network was smaller but remained statistically significant ($\Delta AUC = 0.016$, $p = 0.038$). Bootstrap analysis using 1,000 resamples produced a 95% confidence interval of 0.951–0.982 for the AUC of the explainable model.

McNemar analysis of paired classification outcomes also demonstrated a significant reduction in overall diagnostic errors when the explainable model was compared with the conventional deep neural network ($\chi^2 = 5.21$, $p = 0.022$). The deep neural network incorrectly classified 35 of the 372 independent testing samples, whereas the explainable model produced 29 incorrect classifications. Sensitivity analysis using alternative feature-selection thresholds generated AUC values ranging from 0.959 to 0.970, suggesting that the principal findings were not dependent on a single feature-selection configuration. Repeated stratified cross-validation produced a mean accuracy of 0.919 ± 0.011, supporting the consistency of the observed diagnostic improvement across different partitions of the data.

A positive relationship was observed between explanation stability and diagnostic reliability across the evaluated genomic samples. Predictions associated with highly stable explanatory patterns had a mean classification accuracy of 95.1%, whereas predictions with low explanation stability achieved an accuracy of 84.7%. The correlation between local explanation stability and prediction confidence was moderate and positive ($r = 0.58$, $p < 0.001$), indicating that predictions supported by consistently influential genomic features were generally accompanied by greater model certainty. Samples characterized by highly fragmented attribution patterns were more likely to fall near the classification threshold and were disproportionately represented among false-positive and false-negative cases.

Biological feature coherence was also associated with diagnostic performance. Samples in which the most influential genomic features mapped to at least two functionally related biological pathways achieved an accuracy of 94.8%, compared with 86.2% among samples whose explanatory features showed limited pathway coherence. A multivariable logistic analysis indicated that explanation stability remained independently associated with correct classification after adjustment for age, sex, missingness, and overall genomic signal intensity (odds ratio = 2.34, 95% CI: 1.52–3.61, $p < 0.001$). These relationships suggest that

interpretability metrics provided information beyond conventional prediction probabilities and could be used to identify cases requiring greater diagnostic caution.

Case-level evaluation provided a more detailed view of how the model integrated multiple genomic signals to generate individual diagnostic predictions. One representative disease-positive sample received a predicted disease probability of 0.96. The strongest positive contributions arose from elevated expression of three disease-associated molecular features and a genomic variant involved in a relevant signaling pathway. These variables collectively accounted for 61.3% of the local attribution magnitude. Two protective molecular features reduced the predicted probability by approximately 0.09, but their combined influence was insufficient to alter the final classification. The model prediction was consistent with the recorded diagnostic label and remained stable across all repeated perturbation analyses.

A more ambiguous case involved a disease-positive sample that received an initial predicted probability of 0.57, placing it close to the decision threshold. Local explanation analysis revealed competing genomic signals, with inflammatory pathway activation increasing the predicted disease probability while several metabolic features shifted the prediction toward the reference class. The explanation stability score for this sample was 0.61, substantially lower than the cohort median of 0.82. Removal of a single high-influence genomic feature reduced the predicted probability to 0.48, demonstrating considerable sensitivity to the local molecular configuration. This case illustrates how explainability can identify diagnostically uncertain observations that might otherwise be presented as straightforward classifications.

The global explainability analysis indicated that diagnostic predictions were generated through distributed genomic interactions rather than dependence on one dominant biomarker. Individual high-ranking predictors contributed important information, yet interaction analysis demonstrated that combinations of molecular variables frequently produced larger changes in predicted risk than isolated features. The strongest pairwise interactions were observed between genomic variables associated with cell proliferation and immune-response pathways. The average interaction contribution among the 10 highest-ranked feature pairs was 1.7 times greater in disease-positive samples than in reference samples, suggesting that nonlinear relationships among molecular processes played an important role in diagnostic differentiation.

Error-focused explanation analysis provided additional evidence regarding model behavior. False-positive predictions were commonly associated with strong inflammatory or stress-response genomic signatures that resembled disease-related patterns despite the absence of a positive diagnostic label. False-negative predictions were more frequently found among samples with weak genomic signal intensity, heterogeneous molecular characteristics, or conflicting pathway-level evidence. Approximately 69% of all misclassified samples had explanation stability scores below the cohort median. The concentration of diagnostic errors among biologically ambiguous cases indicates that the explainability framework was useful not only for interpreting successful predictions but also for revealing why the model failed in particular observations.

The findings demonstrate that integrating explainability into artificial intelligence model development can improve diagnostic performance without sacrificing the capacity to process complex, high-dimensional genomic information. The proposed explainable model achieved the highest accuracy, sensitivity, specificity, F1-score, and AUC among the evaluated algorithms. The magnitude of improvement over the conventional deep neural network was relatively modest compared with the differences observed against simpler benchmark methods, which indicates that explainability should not be interpreted as producing a dramatic improvement in every aspect of prediction. Its principal contribution was the combination of strong predictive performance with stable and biologically interpretable evidence supporting individual and population-level classifications.

The results also show that explanation quality has practical relevance for evaluating diagnostic confidence. Stable feature-attribution patterns were associated with higher

classification accuracy, whereas unstable and biologically conflicting explanations were more common among incorrect or borderline predictions. These findings suggest that diagnostic systems should not rely exclusively on predicted probability or binary classification outputs when complex genomic data are involved. A model capable of reporting both diagnostic predictions and the biological evidence underlying those predictions offers a more defensible framework for genomic research. The present results therefore support the use of explainable artificial intelligence as a complementary mechanism for improving diagnostic transparency, identifying uncertain cases, and strengthening the biological interpretability of computational decision-making.

The present findings demonstrate that the proposed explainable artificial intelligence framework achieved the strongest diagnostic performance among the evaluated computational models while preserving interpretable information about the genomic features contributing to individual predictions. The model reached an accuracy of 0.923, sensitivity of 0.931, specificity of 0.916, F1-score of 0.925, and an AUC of 0.968 on the independent testing dataset. These values exceeded the performance of logistic regression, support vector machine, random forest, extreme gradient boosting, and the conventional deep neural network. The advantage over simpler machine-learning approaches was substantial, whereas the improvement over the deep neural network was comparatively modest. This pattern indicates that the main value of the proposed framework should not be reduced to a numerical increase in classification performance. Its more important contribution lies in combining strong diagnostic discrimination with interpretable evidence about how genomic information influenced the prediction process.

The comparative results also reveal that increasingly flexible models were generally better able to discriminate between disease and reference samples. Logistic regression produced the lowest AUC of 0.881, while progressively more complex algorithms achieved higher values, culminating in an AUC of 0.952 for the conventional deep neural network and 0.968 for the proposed explainable model. The pattern suggests that nonlinear relationships among genomic variables contained diagnostically useful information that could not be fully captured by simpler linear decision boundaries. The improvement in calibration further strengthens this interpretation because the explainable model produced a lower Brier score than the conventional neural network. Diagnostic usefulness depends not only on whether a model assigns the correct class but also on whether its estimated probabilities correspond reasonably well to observed outcomes.

The explainability analysis provided a second major finding by showing that diagnostic decisions were concentrated around a relatively small but stable group of genomic variables. Twenty-five features accounted for a substantial proportion of the total attribution magnitude, while 18 of these features remained consistently highly ranked across repeated validation procedures. The mean Spearman correlation of 0.84 for global feature rankings indicates that the model was not simply producing a different explanatory structure whenever the sample composition changed. Local explanations were less stable than global explanations, which is understandable because patient-level predictions depend on specific combinations of genomic signals. This distinction is important because model-level interpretability and individual-level interpretability address different scientific questions and should not be treated as interchangeable measures.

The error analysis adds further meaning to the overall results. Incorrect classifications were disproportionately associated with low explanation stability, conflicting molecular patterns, and predictions close to the classification threshold. Samples with highly stable explanations achieved greater diagnostic accuracy than samples with unstable attribution patterns, while the representative borderline case showed that a single influential feature could materially alter the predicted diagnosis (A. K. Singh et al., 2025; Zafar et al., 2023). Such findings suggest that explainability may serve not merely as a descriptive layer placed over a

trained model but as an indicator of diagnostic uncertainty. A probability score may indicate how confident the model appears to be, whereas explanation stability can reveal whether that confidence is supported by a consistent underlying genomic pattern.

The results are broadly consistent with the expanding literature showing that explainable artificial intelligence can combine predictive performance with greater transparency in biomedical and precision-medicine applications. A systematic review of XAI in disease prediction reported that SHAP and LIME remain among the most commonly used explanation techniques and emphasized their growing application to genetic, physiological, imaging, and other medical data. The same review identified limited dataset diversity, dependence on single data modalities, and difficulties in clinician interpretation as persistent weaknesses in the literature. The present findings extend this discussion by treating explanation stability and biological coherence as part of model evaluation rather than presenting feature-importance plots as sufficient evidence of interpretability. This distinction is important because an explanation that appears intuitive is not necessarily reproducible or scientifically meaningful.

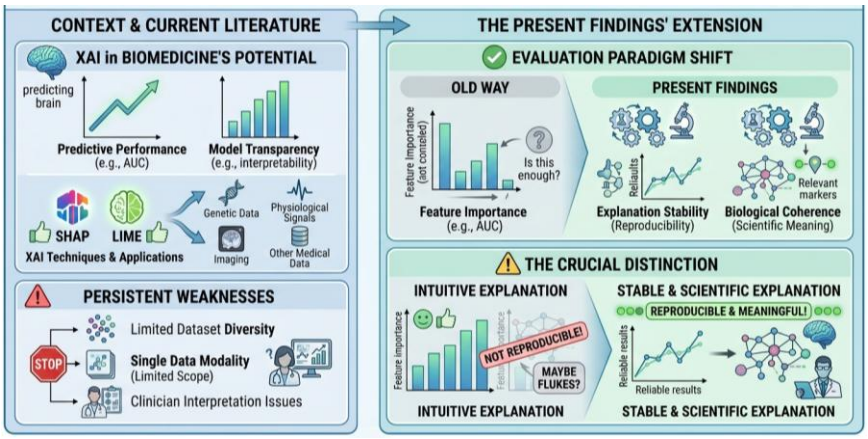


Figure 2. Translating XAI Concepts in Biomedicine: Beyond Feature Importance Plots

Recent genomic research provides direct support for the potential value of biologically interpretable artificial intelligence. Tripathy and colleagues developed an explainable graph neural network integrating multi-omics information with prior biological knowledge for biomarker identification, demonstrating the value of linking computational learning with known biological relationships. Research using tumor-educated platelet RNA similarly showed that SHAP-based interpretation could identify consistent gene-level contributors and context-dependent interactions while maintaining strong predictive performance across multiple cancer types. Interpretable multi-omics architectures have also achieved high classification performance across multiple cancer categories while preserving feature-level biological information. The present results align with this direction by showing that interpretability and predictive capability do not have to be treated as mutually exclusive objectives.

A meaningful difference emerges when these findings are compared with studies showing that additional molecular complexity does not automatically produce superior diagnostic models. Research using large-scale UK Biobank data found that multi-omics biomarkers did not consistently outperform conventional clinical variables for pan-cancer screening, although stronger performance was observed for several individual cancer types. This evidence challenges the assumption that larger numbers of genomic or molecular variables inherently generate better predictions. The current findings should therefore be interpreted as evidence for the value of selectively extracting informative biological patterns rather than evidence that genomic complexity itself guarantees diagnostic superiority. Effective feature selection, model regularization, data quality, disease context, and biological relevance remain central determinants of performance (Nayyar et al., 2025; Rozera et al., 2025).

The present study also responds to concerns identified in recent evaluations of XAI research. A systematic review of explainable AI in cancer diagnosis, prognosis, and treatment planning reported that a large proportion of published studies did not involve clinician input and that most failed to formally evaluate the quality of their explanations. The authors specifically cautioned that post-hoc approaches such as SHAP and LIME can produce misleading interpretations when their assumptions or limitations are ignored. A broader systematic review of XAI in precision medicine similarly concluded that further development and validation of explanation methods are necessary before explainability can reliably support clinical decision-making. The emphasis placed here on explanation stability, fidelity, biological plausibility, and error analysis addresses part of that methodological weakness, although it does not eliminate the need for external and clinical validation.

The findings signal an important shift in the criteria used to judge artificial intelligence models for genomic diagnostics. Predictive accuracy has traditionally served as the dominant benchmark, especially when machine-learning models are compared using metrics such as accuracy, sensitivity, specificity, and AUC. High-stakes biomedical applications demand a broader conception of model quality. A system that produces accurate classifications for reasons that cannot be examined, replicated, or connected to credible biological evidence remains scientifically incomplete. The present results suggest that diagnostic performance and interpretability should be evaluated as complementary properties. This perspective moves the field from asking whether an algorithm predicts correctly toward asking whether it predicts correctly for reasons that can be critically inspected.

The stability of genomic explanations is particularly significant because it provides information about the reproducibility of model reasoning. A genomic feature that repeatedly appears among the dominant predictors across resampled datasets has greater analytical credibility than a feature that becomes important only under one specific model initialization or sampling configuration (Górriz et al., 2023; Noviandy et al., 2025). Such stability should still not be interpreted as proof of biological causality. SHAP values, permutation importance, and related attribution measures explain the behavior of a predictive model; they do not by themselves establish that a gene or molecular pathway causes a disease phenotype. The distinction prevents a common interpretive error in XAI research, where computational importance is sometimes presented too quickly as biological mechanism.

The concentration of unstable explanations among false-positive and false-negative predictions may also be interpreted as evidence of underlying biological heterogeneity. Complex diseases rarely arise from one uniform molecular mechanism, and patients assigned to the same diagnostic category may exhibit substantially different genomic profiles. Borderline predictions may therefore reflect more than algorithmic weakness. Some may represent atypical disease subtypes, incomplete phenotypic characterization, overlapping molecular pathways, or cases situated biologically between conventional diagnostic categories. Explainability can help reveal this heterogeneity by showing when a classification depends on competing molecular signals rather than a coherent diagnostic signature.

The broader significance of these findings concerns the relationship between artificial intelligence and precision medicine. Clinical genetics increasingly incorporates AI into variant interpretation, phenotype analysis, risk estimation, and other genomic tasks, while questions concerning generalizability, interpretability, and responsible use remain unresolved (Ben Achour et al., 2025; Seraphin et al., 2025). The present findings suggest that explainable models may help move genomic diagnostics away from population-level classification alone toward more individualized assessments in which both the predicted outcome and its molecular basis can be examined. Such a transition would be consistent with the goals of precision medicine, but genuine personalization requires validation across populations and healthcare environments rather than merely producing patient-specific explanation plots.

The immediate clinical implication is that explainable artificial intelligence could support a more cautious form of AI-assisted diagnostic decision-making. The proposed framework should not be interpreted as a replacement for clinicians, geneticists, pathologists, or laboratory specialists. Its stronger role would be to provide a structured second layer of evidence: a diagnostic probability accompanied by information about the genomic features that drove the prediction and the stability of those explanations. Clinicians could then distinguish between predictions supported by coherent molecular evidence and predictions generated from unstable or contradictory patterns. Such a system would make algorithmic uncertainty more visible rather than concealing it behind a single binary output.

The implications for genomic biomarker discovery are equally important. Stable high-ranking features and recurrent feature interactions may help researchers prioritize genomic candidates for functional investigation. Recent studies using SHAP and explainable genomic models have demonstrated that interpretable machine learning can identify genes associated with cancer classification, prognosis, and regulatory interactions (Baheti et al., 2025; Holzinger et al., 2023). The present findings reinforce this potential but also impose an important qualification: computational attribution should be treated as a hypothesis-generating mechanism. Candidate genes identified through XAI require confirmation through independent cohorts, pathway analysis, experimental assays, and where appropriate functional studies before they can be considered validated biomarkers.

The methodological implication concerns how future diagnostic AI studies should report model quality. Accuracy and AUC alone provide an incomplete account of performance in biologically complex datasets. Calibration, subgroup performance, explanation stability, robustness to perturbation, feature consistency, and biological plausibility should become routine components of evaluation. The association observed between explanation stability and correct classification indicates that interpretability metrics may contain diagnostic information that conventional performance statistics overlook. A standardized evaluation framework combining prediction and explanation metrics could make comparisons between XAI studies more meaningful and reduce the tendency to label a model “explainable” simply because a SHAP summary plot or similar visualization has been added.

The translational implication extends to data governance and population representation. Genomic models can inherit population stratification, sequencing biases, laboratory effects, and uneven representation within training datasets. A highly accurate model developed predominantly from one ancestry group, institution, or sequencing platform may perform less reliably elsewhere. Precision medicine itself carries risks of widening health disparities when advanced genomic technologies are not validated or made accessible across diverse populations. Explainability provides an opportunity to examine whether different groups are classified using similar biological evidence, but explanation alone cannot remove dataset bias. Subgroup validation and representative sampling remain indispensable.

The superior performance of the explainable model can first be understood in relation to the structure of genomic data. Biological outcomes are produced through interacting molecular systems rather than isolated predictors, while gene expression, genomic variants, regulatory processes, and clinical factors frequently exhibit nonlinear and context-dependent relationships. Linear algorithms may fail to represent such dependencies adequately. Deep-learning and ensemble approaches can capture more complex feature interactions, which helps explain why the simpler benchmark models produced lower diagnostic performance. The relatively small difference between the deep neural network and the proposed explainable framework also indicates that much of the predictive gain originated from nonlinear representation learning rather than explainability itself.

The additional improvement associated with the explainable framework may be attributable to the way interpretability was integrated with feature refinement and model evaluation. High-dimensional genomic datasets typically contain thousands of variables, many

of which are redundant, noisy, weakly informative, or vulnerable to technical variation. A model that learns indiscriminately from this feature space can achieve strong training performance while relying on unstable associations. Interpretability-guided evaluation makes it possible to identify variables that repeatedly dominate predictions and to examine whether those variables remain consistent across validation conditions. Such a process can reduce dependence on incidental correlations and focus learning on more reproducible combinations of biological information (Haferlach et al., 2025; Santorsola & Lescai, 2023).

The positive association between explanation stability and classification accuracy can also be interpreted through signal coherence. Samples with clear disease-related molecular signatures are likely to produce several mutually reinforcing genomic signals, allowing the model to reach similar explanations when the data or model parameters are slightly perturbed. Samples with heterogeneous, weak, or contradictory molecular signals are more likely to produce unstable feature rankings because multiple alternative combinations can lead to similar prediction probabilities. Explanation instability may therefore function as a computational expression of biological ambiguity. This mechanism would account for the higher frequency of low-stability explanations among misclassified and near-threshold samples.

The remaining diagnostic errors are likely to reflect a combination of biological and methodological factors rather than one single limitation. Disease heterogeneity, imperfect diagnostic labels, sequencing noise, missing molecular information, population structure, batch effects, rare subtypes, and unmeasured environmental factors can all weaken the relationship between genomic predictors and observed diagnosis. Biological systems also contain redundancy, meaning that different molecular pathways may produce similar phenotypes. Artificial intelligence can model many of these relationships, but it cannot recover relevant information that is absent from the dataset. The persistence of false-positive and false-negative cases therefore reinforces the need to interpret even high-performing genomic AI as probabilistic evidence rather than as an infallible diagnostic authority.

The next research priority should be rigorous external validation across independent institutions, populations, sequencing technologies, and disease contexts. Internal train-validation-test separation provides protection against straightforward overfitting but cannot establish generalizability to settings that differ materially from the development dataset. Multi-center validation should examine whether predictive accuracy, calibration, feature rankings, and explanation stability remain consistent when the model encounters new population structures and laboratory conditions. Particular attention should be directed toward ancestry-related genomic variation because models developed from poorly representative datasets may unintentionally produce unequal diagnostic performance across demographic groups.

Biological validation should constitute a second priority. Stable XAI features should be compared with established disease genes, pathway databases, protein-interaction networks, and independent transcriptomic or genomic cohorts. Experimental validation using cellular, molecular, or functional assays would be required for high-priority candidate biomarkers before causal claims could be justified. Explainable graph-based approaches that incorporate prior biological knowledge provide one promising direction because they can model genes as components of interacting biological systems rather than independent numerical variables. Recent work integrating multi-omics data with prior knowledge through explainable graph neural networks illustrates the potential of this approach for biomarker discovery.

Prospective clinical evaluation should then determine whether explanations actually improve human decision-making. A model can produce technically faithful explanations without making clinicians more accurate, faster, or appropriately cautious. Recent evidence concerning human interaction with XAI indicates that clinician trust and reliance can vary considerably between individuals, highlighting the importance of studying the human factors surrounding explainable systems rather than assuming that more explanation automatically leads to better decisions. Future trials should therefore compare diagnostic decisions made

without AI assistance, with opaque AI assistance, and with explainable AI assistance. Outcomes should include diagnostic accuracy, time to decision, confidence calibration, inappropriate reliance on incorrect predictions, and clinician comprehension of genomic explanations.

Future technical development should move toward multimodal and biologically structured models that integrate genomic, transcriptomic, epigenomic, proteomic, imaging, and clinical information while explicitly representing uncertainty. Current reviews identify data integration and dataset diversity as major unresolved challenges for XAI-based disease prediction. Graph neural networks, attention-based architectures, biologically informed representation learning, and emerging foundation models may offer stronger ways to capture relationships across biological scales, but greater architectural sophistication should not become an end in itself. The next generation of genomic AI should be judged by whether it produces predictions that remain accurate across settings, exposes uncertainty honestly, generates explanations that are reproducible and biologically defensible, and ultimately improves clinical or scientific decisions. Explainability becomes valuable when it strengthens the evidentiary chain between complex biological data, computational prediction, scientific interpretation, and responsible diagnostic action.

CONCLUSION

The most important finding of this study is that explainable artificial intelligence can improve diagnostic performance while simultaneously providing a more transparent account of how complex genomic features contribute to model predictions. The proposed model achieved stronger overall performance than the conventional machine-learning and deep-learning benchmarks, particularly in accuracy, sensitivity, specificity, F1-score, and area under the receiver operating characteristic curve. Explanation stability was also associated with higher classification reliability, whereas unstable genomic attribution patterns were more frequently observed in borderline or misclassified cases. This finding distinguishes the study from approaches that evaluate explainability only as a visualization tool, because the results indicate that interpretability can also provide useful information about prediction confidence, biological coherence, and potential diagnostic uncertainty.

The principal contribution of this research is both methodological and conceptual. Methodologically, the study integrates predictive modeling, feature-attribution analysis, explanation stability assessment, biological plausibility evaluation, and diagnostic error analysis within a single explainable artificial intelligence framework. This approach extends beyond conventional post-hoc interpretation by positioning explainability as an active component of model evaluation and refinement. Conceptually, the study proposes that the quality of genomic diagnostic models should not be judged solely by predictive accuracy, but also by the consistency, reproducibility, and biological interpretability of the evidence underlying their predictions. This perspective contributes to the development of more trustworthy artificial intelligence systems for genomic research and supports a more rigorous connection between computational prediction and biomedical interpretation.

The study remains limited by its retrospective computational design, dependence on the characteristics and quality of the available genomic dataset, and the absence of prospective clinical and laboratory validation. High-dimensional genomic data are vulnerable to population stratification, technical variation, missing information, and disease heterogeneity, all of which may influence both predictive performance and explanatory stability. Feature importance generated by explainable artificial intelligence should therefore not be interpreted as direct evidence of biological causality. Future studies should validate the proposed framework across independent multi-center cohorts, more diverse ancestry groups, different sequencing platforms, and additional disease categories. Further research should also integrate multi-omics and clinical data, conduct experimental validation of high-priority genomic features, compare

alternative explainability techniques, and evaluate whether interpretable model outputs improve real-world clinical decision-making, diagnostic confidence, and patient-level outcomes.

DECLARATION OF AI AND AI ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the author(s) used ChatGPT and QuillBot solely to assist with text translation. After using these tools/services, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., Albahri, O. S., Alamoodi, A. H., Bai, J., Salhi, A., Santamaría, J., Ouyang, C., Gupta, A., Gu, Y., & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. Information Fusion, 96, 156–191. https://doi.org/https://doi.org/10.1016/j.inffus.2023.03.008

Bae, H., Park, S.-Y., & Kim, C.-E. (2024). A practical guide to implementing artificial intelligence in traditional East Asian medicine research. Integrative Medicine Research, 13(3), 101067. https://doi.org/https://doi.org/10.1016/j.imr.2024.101067

Baheti, B., Rai, S., Innani, S., Mehdiratta, G., Bell, W. R., Guntuku, S. C., Nasrallah, M. P., & Bakas, S. (2025). Multimodal Explainable Artificial Intelligence for Prognostic Stratification of Patients With Glioblastoma. Modern Pathology, 38(9), 100797. https://doi.org/https://doi.org/10.1016/j.modpat.2025.100797

Bakrania, A., Joshi, N., Zhao, X., Zheng, G., & Bhat, M. (2023). Artificial intelligence in liver cancers: Decoding the impact of machine learning models in clinical diagnosis of primary liver cancers and liver cancer metastases. Pharmacological Research, 189, 106706. https://doi.org/https://doi.org/10.1016/j.phrs.2023.106706

Ben Achour, T., Boucherb, M., Jridi, M., Naceur, I., Ben Ghorbel, I., Smiti, M., & Said, F. (2025). Development of an artificial intelligence model for predicting the etiology of venous thromboembolism. Intelligence-Based Medicine, 12, 100284. https://doi.org/https://doi.org/10.1016/j.ibmed.2025.100284

Cilluffo, G., Bajowala, S., Elliott, T., Gupta, P., Volertas, S., & Ferrante, G. (2025). The Future of Allergy Management: How Artificial Intelligence Is Changing the Game. The Journal of Allergy and Clinical Immunology: In Practice, 13(11), 2877–2889. https://doi.org/https://doi.org/10.1016/j.jaip.2025.08.033

Croce, D., Smirnov, A., Tiburzi, L., Travaglini, S., Costa, R., Calabrese, A., Basili, R., Levialdi Ghiron, N., & Melino, G. (2024). AI-driven transcriptomic encoders: From explainable models to accurate, sample-independent cancer diagnostics. Expert Systems with Applications, 258, 125126. https://doi.org/https://doi.org/10.1016/j.eswa.2024.125126

D'Ascenzo, F., Rocchi, A., Iandolo, F., & Vito, P. (2024). Evolutionary impacts of artificial intelligence in healthcare managerial literature. A ten-year bibliometric and topic

modeling review. Sustainable Futures, 7, 100198. https://doi.org/https://doi.org/10.1016/j.sftr.2024.100198

de Oliveira Avellar, W., Ferreira, É. A., & Aran, V. (2024). Artificial Intelligence and cancer: Profile of registered clinical trials. Journal of Cancer Policy, 42, 100503. https://doi.org/https://doi.org/10.1016/j.jcpo.2024.100503

Eti, S., Yüksel, S., Eti, S. T., Dinçer, H., & Eyupoglu, O. E. (2025). Leveraging artificial intelligence and koch snowflake fuzzy sets to optimize antibiotic development pathways. Artificial Intelligence in the Life Sciences, 8, 100144. https://doi.org/https://doi.org/10.1016/j.ailsci.2025.100144

Gangas, P., Graf, N., Wen, S., Cattaneo, C., Bicchieri, M., & Lennox-Chhugani, N. (2025). Artificial intelligence in paediatric cancer: Insights from innovation experts in the UNICA4EU project. EJC Paediatric Oncology, 5, 100208. https://doi.org/https://doi.org/10.1016/j.ejcped.2024.100208

Gogali, A., Kyriakopoulos, C., Potonos, D., Ladias, A., Tatsis, K., & Kostikas, K. (2025). Artificial Intelligence in Asthma: Current Status, Opportunities, and Pitfalls. The Journal of Allergy and Clinical Immunology: In Practice, 13(11), 2891–2901. https://doi.org/https://doi.org/10.1016/j.jaip.2025.05.044

Górriz, J. M., Álvarez-Illán, I., Álvarez-Marquina, A., Arco, J. E., Atzmueller, M., Ballarini, F., Barakova, E., Bologna, G., Bonomini, P., Castellanos-Dominguez, G., Castillo-Barnes, D., Cho, S. B., Contreras, R., Cuadra, J. M., Domínguez, E., Domínguez-Mateos, F., Duro, R. J., Elizondo, D., Fernández-Caballero, A., … Ferrández-Vicente, J. M. (2023). Computational approaches to Explainable Artificial Intelligence: Advances in theory, applications and trends. Information Fusion, 100, 101945. https://doi.org/https://doi.org/10.1016/j.inffus.2023.101945

Guo, K., Wu, M., Soo, Z., Yang, Y., Zhang, Y., Zhang, Q., Lin, H., Grosser, M., Venter, D., Zhang, G., & Lu, J. (2023). Artificial intelligence-driven biomedical genomics. Knowledge-Based Systems, 279, 110937. https://doi.org/https://doi.org/10.1016/j.knosys.2023.110937

Gutierrez-Chakraborty, E., Chakraborty, D., Das, D., & Bai, Y. (2024). Discovering novel prognostic biomarkers of hepatocellular carcinoma using eXplainable Artificial Intelligence. Expert Systems with Applications, 252, 124239. https://doi.org/https://doi.org/10.1016/j.eswa.2024.124239

Haferlach, T., Eckardt, J.-N., Walter, W., Maschek, S., Kather, J. N., Pohlkamp, C., & Middeke, J. M. (2025). AML diagnostics in the 21st century: Use of AI. Seminars in Hematology, 62(3), 226–234. https://doi.org/https://doi.org/10.1053/j.seminhematol.2025.06.002

Hernández-Peñaloza, G., Uribe, S., García, F. M., Graf, N., & Álvarez, F. (2024). General context and relevant public datasets available for improving pathways in Paediatric Cancer applying Artificial Intelligence. A review. EJC Paediatric Oncology, 4, 100196. https://doi.org/https://doi.org/10.1016/j.ejcped.2024.100196

Holzinger, A., Keiblinger, K., Holub, P., Zatloukal, K., & Müller, H. (2023). AI for life: Trends in artificial intelligence for biotechnology. New Biotechnology, 74, 16–24. https://doi.org/https://doi.org/10.1016/j.nbt.2023.02.001

Hurkmans, C., Bibault, J.-E., Brock, K. K., van Elmpt, W., Feng, M., David Fuller, C., Jereczek-Fossa, B. A., Korreman, S., Landry, G., Madesta, F., Mayo, C., McWilliam, A., Moura, F., Muren, L. P., El Naqa, I., Seuntjens, J., Valentini, V., & Velec, M. (2024). A joint ESTRO and AAPM guideline for development, clinical validation and reporting of artificial intelligence models in radiation therapy. Radiotherapy and Oncology, 197, 110345. https://doi.org/https://doi.org/10.1016/j.radonc.2024.110345

Iacucci, M., Santacroce, G., Yasuharu, M., & Ghosh, S. (2025). Artificial Intelligence–Driven Personalized Medicine: Transforming Clinical Practice in Inflammatory Bowel Disease.

Gastroenterology, 169(3), 416–431. https://doi.org/https://doi.org/10.1053/j.gastro.2025.03.005

Kamphorst, K., de Jong, J., Rider, N. L., & Portnoy, J. M. (2025). Big Data and Artificial Intelligence: Current State and Future Opportunities in Allergy and Immunology. The Journal of Allergy and Clinical Immunology: In Practice, 13(11), 2914–2924. https://doi.org/https://doi.org/10.1016/j.jaip.2025.09.011

Kaniwa, F. (2025). GenRepAI: Utilizing Artificial Intelligence to Identify Repeats in Genomic Suffix Trees. Current Bioinformatics, 20(6), 522–534. https://doi.org/https://doi.org/10.2174/0115748936303435240702112205

Li, J., Han, X., Qin, Y., Tan, F., Chen, Y., Wang, Z., Song, H., Zhou, X., Zhang, Y., Hu, L., & Hu, P. (2023). Artificial intelligence accelerates multi-modal biomedical process: A Survey. Neurocomputing, 558, 126720. https://doi.org/https://doi.org/10.1016/j.neucom.2023.126720

Lin, S.-Y., Tsai, P.-C., Su, F.-Y., Chen, C.-Y., Li, F., Zhao, J., Ho, Y. Y., Lee, T.-L. M., Healey, E., Lin, P.-J., Kao, T.-W., Vremenko, D., Roetzer-Pejrimovsky, T., Sholl, L., Dillon, D., Lin, N. U., Meredith, D., Ligon, K. L., Lo, Y.-C., … Yu, K.-H. (2025). Contrastive learning enhances fairness in pathology artificial intelligence systems. Cell Reports Medicine, 6(12), 102527. https://doi.org/https://doi.org/10.1016/j.xcrm.2025.102527

Maji, A. K., Das, S., Marwaha, S., Kumar, S., Dutta, S., Choudhury, M. R., Arora, A., Ray, M., Perumal, A., & Chinusamy, V. (2025). Intelligent decision support for drought stress (IDSDS): An integrated remote sensing and artificial intelligence-based pipeline for quantifying drought stress in plants. Computers and Electronics in Agriculture, 236, 110477. https://doi.org/https://doi.org/10.1016/j.compag.2025.110477

Nayyar, A., Shrivastava, R., & Jain, S. (2025). A Framework for Two-class Classification of Pulmonary Tuberculosis using Artificial Intelligence. Current Medical Imaging, 21. https://doi.org/https://doi.org/10.2174/0115734056343819250307040956

Noviandy, T. R., Idroes, G. M., & Hardi, I. (2025). Integrating explainable artificial intelligence and light gradient boosting machine for glioma grading. Informatics and Health, 2(1), 1–8. https://doi.org/https://doi.org/10.1016/j.infoh.2024.12.001

Reyes Gil, M., Pantanowitz, J., & Rashidi, H. H. (2024). Venous thromboembolism in the era of machine learning and artificial intelligence in medicine. Thrombosis Research, 242, 109121. https://doi.org/https://doi.org/10.1016/j.thromres.2024.109121

Rozera, T., Pasolli, E., Segata, N., & Ianiro, G. (2025). Machine Learning and Artificial Intelligence in the Multi-Omics Approach to Gut Microbiota. Gastroenterology, 169(3), 487–501. https://doi.org/https://doi.org/10.1053/j.gastro.2025.02.035

Santorsola, M., & Lescai, F. (2023). The promise of explainable deep learning for omics data analysis: Adding new discovery tools to AI. New Biotechnology, 77, 1–11. https://doi.org/https://doi.org/10.1016/j.nbt.2023.06.002

Seraphin, T. P., Mesropian, A., Žigutytė, L., Brooks, J., Mauro, E., Gris-Oliver, A., Pinyol, R., Montironi, C., Balaseviciute, U., Piqué-Gili, M., Huguet-Pradell, J., van Treeck, M., Kallenbach, M., Schneider, A. T., Roderburg, C., Kather, J. N., Luedde, T., & Llovet, J. M. (2025). Artificial intelligence predicts outcome-related molecular profiles and vascular invasion in hepatocellular carcinoma. JHEP Reports, 7(12), 101592. https://doi.org/https://doi.org/10.1016/j.jhepr.2025.101592

Singh, A. (2024). Artificial intelligence for drug repurposing against infectious diseases. Artificial Intelligence Chemistry, 2(2), 100071. https://doi.org/https://doi.org/10.1016/j.aichem.2024.100071

Singh, A. K., Jones, S. E., Van der Laan, L., Ayanlade, T. T., Raigne, J., Saleem, N., Joshi, S., Arshad, M. A., ZareMehrjerdi, H., Rairdin, A., Di Salvo, J., Elango, D., De Azevedo Peixoto, L., Jubery, T. Z., Krishnamurthy, A., Singh, A., Sarkar, S., &

Ganapathysubramanian, B. (2025). Chapter Five - Use of artificial intelligence in soybean breeding and production (D. L. B. T.-A. in A. Sparks (ed.); Vol. 190, pp. 199–273). Academic Press. https://doi.org/https://doi.org/10.1016/bs.agron.2024.11.003

Singh, S. S., Kumar, S., Ahuja, R., & Barua, J. (2025). Fusion of quantum computing and explainable AI: A comprehensive survey on transformative healthcare solutions. Information Fusion, 122, 103217. https://doi.org/https://doi.org/10.1016/j.inffus.2025.103217

Sun, D., Li, H., Wang, Y., Li, D., Xu, D., & Zhang, Z. (2024). Artificial intelligence-based pathological application to predict regional lymph node metastasis in Papillary Thyroid Cancer. Current Problems in Cancer, 53, 101150. https://doi.org/https://doi.org/10.1016/j.currproblcancer.2024.101150

Torres-Martos, Á., Anguita-Ruiz, A., Bustos-Aibar, M., Ramírez-Mena, A., Arteaga, M., Bueno, G., Leis, R., Aguilera, C. M., Alcalá, R., & Alcalá-Fdez, J. (2024). Multiomics and eXplainable artificial intelligence for decision support in insulin resistance early diagnosis: A pediatric population-based longitudinal study. Artificial Intelligence in Medicine, 156, 102962. https://doi.org/https://doi.org/10.1016/j.artmed.2024.102962

Wei, W., Song, C., Qi, C., Zhang, X., Zhang, X., Pu, R., & Ao, Y. (2024). Visual analysis of drug research and development based on artificial intelligence. Journal of Holistic Integrative Pharmacy, 5(4), 323–332. https://doi.org/https://doi.org/10.1016/j.jhip.2024.12.002

Wijaya, J., Park, J., Yang, Y., Siddiqui, S. I., & Oh, S. (2024). A metagenome-derived artificial intelligence modeling framework advances the predictive diagnosis and interpretation of petroleum-polluted groundwater. Journal of Hazardous Materials, 472, 134513. https://doi.org/https://doi.org/10.1016/j.jhazmat.2024.134513

Wu, P., Chen, L., Xu, Y.-X., Ren, Y.-Y., Dong, X.-M., Fan, J.-X., Xie, T., Zhou, J.-L., & Liu, Z.-Q. (2025). Decoding herbal medicine: AI-powered omics and network pharmacology. Phytomedicine, 148, 157453. https://doi.org/https://doi.org/10.1016/j.phymed.2025.157453

Wu, Y., Wang, K., Song, Y., & Li, B. (2025). Enhancing ovarian cancer prognosis with an artificial intelligence-derived model: Multi-omics integration and therapeutic implications. Translational Oncology, 59, 102439. https://doi.org/https://doi.org/10.1016/j.tranon.2025.102439

Xin, X., Wu, S., Xu, H., Ma, Y., Bao, N., Gao, M., Han, X., Gao, S., Zhang, S., Zhao, X., Qi, J., Zhang, X., & Tan, J. (2024). Non-invasive prediction of human embryonic ploidy using artificial intelligence: a systematic review and meta-analysis. EClinicalMedicine, 77, 102897. https://doi.org/https://doi.org/10.1016/j.eclinm.2024.102897

Zafar, I., Anwar, S., kanwal, F., Yousaf, W., Un Nisa, F., Kausar, T., ul Ain, Q., Unar, A., Kamal, M. A., Rashid, S., Khan, K. A., & Sharma, R. (2023). Reviewing methods of deep learning for intelligent healthcare systems in genomics and biomedicine. Biomedical Signal Processing and Control, 86, 105263. https://doi.org/https://doi.org/10.1016/j.bspc.2023.105263

Copyright Holder :
© Mohammad Ma'arif Ikhsanudin et al. (2026).

First Publication Right :
© Research of Scientia Naturalis

This article is under:

