GENETIC RISK PREDICTOR FOR CHRONIC ILLNESSES BASED ON FAMILY HISTORY AND LIFESTYLE DATA WITH MEDICAL RECOMMENDATION

1st Mrs. V.Prema Tulasi
Department of CSE (Data Science)
CMR Technical Campus
Hyderabad, India

2nd D.Jamuna
Department of CSE (Data Science)
CMR Technical Campus
Hyderabad, India

3rd V.Guru Mani Chandu
Department of CSE (Data Science)
CMR Technical Campus
Hyderabad, India

4th K.Manasa
Department of CSE (Data Science)
CMR Technical Campus
Hyderabad, India

Abstract-*This research describes an intelligent web-based Genetic Disorder Detection System with Hybrid Machine Learning Inference, Rules based clinical mapping and generative medical guidance. The system uses Decision Tree classifier based on the training on medically relevant features and expanded to 26 diagnostic parameters using engineered mappings of symptom categories and inheritance patterns. With a backend developed in Python using the Flask module, the model takes user inputs, normalises features with the help of the pre-fitted scaler and predicts one of nine genetic disorders with high accuracy. To improve interpretability and clinical application, the system uses the generative model of Google's Gemini which offers personalized explanations, answers to the symptoms, lifestyle recommendations, and genetic counselling advice. User authentication, demographic profiling, real-time prediction workflows etc are enabled with a secure yet nicely simple to operate involves a secure and flexible data architecture, based on a JSON-based database. This integrated approach to this problem allows for accessible, fast screening support of diseases based on genetic predisposition, and bundles deterministic rules, statistical learning and medical instructions in natural language into a single diagnostic aid.*

Keywords*: Genetic disorder detection Machine learning Rule based mapping Medical decision support : Flask web application,Gemini generative AI, Medical informatics Feature engineering, Hybrid diagnostic system*

I. INTRODUCTION

The rapid growth in artificial intelligence in the field of medical genomics has changed the scenario of prediction of diseases and clinical decision making to a great extent. Recent advancements on generative AI technologies - in particular, in the field of human genomics analysis - has proven immense potential for the support of an early diagnosis as well as of personalised healthcare strategies [1]. Generative systems may be useful in synthesising bio-medically relevant insights from complex biological data, in order to deliver more robust models of context-aware predictions for the identification of diseases. Such adoption in terms of disease forecasting is spreading in different areas of clinical practice about the opportunity of greater accuracy and interpretability from the diagnostic process workflows using automated reasoning [2]. As the genomic information databases expand, and the models for computational analysis take off, the fusion of generative AI and classic predictive analytics will be increasingly more important in troubleshooting the characteristics of early detection in genetic-based diseases.

Machine learning has developed into a backbone of precision medicine specifically in the diagnosis of rare and complicated genetic disorder which tend to go undetected by using conventional clinical evaluation methods [3]. Recent literature is blase with the use of combination of analytical frameworks that are knowledge-and data-driven so as to better the performance of diagnostic in heterogeneous population of patients [4]. Furthermore, predictive cellular modelling, with the support of multi-omics integration is seeing potential in enabling a digital representation of the patient biology to support more informed diagnostic reasoning [5]. These advancements in the field highlight the need to create smart diagnostic systems which would be able to tackle genetic multi-factorial conditions, using scalable computational methods.

Concurrently, generative AI is also getting increasing definitive adoption in the bioinformatics, the pharmaceutical sciences and the clinical research fields based on the capability of creating structured intelligence from scattering or incomplete medical data [6][7][8]. A combination of both generative reasoning and rule-based and explainable AI are now being explored in complex disease scenarios including infectious and hereditary disorders [9]. The trend towards agentic and autonomous artificial intelligence systems in health care only serves to reinforce the importance of the transparency, interactive and aligned user diagnostic platforms that can make medical decisions in a responsible manner [10]. In this context the development of a system which would encompass machine learning prediction, rules-based clinical mapping and generative medical explanation in this respect is a promising development in terms of early screening of genetic disorders.

II. LITERATURE SURVEY

Generative Artificial Intelligence and Medical Genomics

Recent surveys and systematic reviews have taken note of the swift rise of the generative AI, as a revolutionarily transforming tool for medical genomics. Variation interpretation and clinical translation challenges Changalidis et al. offer an expansive review addressing how generative architectures can be employed to aid variant interpretation, functional annotation and the augmentation of synthetic data [1]. Complementing this, extensive overviews of generative AI in bioinformatics incorporate methodological advances in design across different model families, training protocols and evaluation metrics and focus on the need for domain-specific restrictions and explainability factors that are necessary to ensure biological implausibility and reliable clinical application [6]. Together these works make the argument that generative models will help speed up genomic discovery, but will need to be carefully governed, have transparent reporting pathways, with clinical validation pathways for use.

Generative Models for Clinical Disease Prediction Multimodal Transformers

Generative models of language and multimodal models are being investigated for clinical support of diagnosis and the prediction of disease. Ghosh illustrates that generative AI can be used to forecast diseases, how the language-model style of reasoning can be used to synthesize clinical information to generate patient-specific diagnostic recommendations [2]. Abugabah et al. propose a knowledge-guided multimodal transformer framework for the fusion of electronic health records and domain knowledge for better diagnostic accuracy for rare diseases, proving to be robust in its ability to work with the heterogeneity of clinical inputs and knowledge constraints [4]. These studies together suggest that designs of hybrid transformers can strengthen the predictive ability, and "improve" the interpretability, of the information they present by combining reasoning in generative way and structured clinical knowledge and EHR data.

Machine Learning for Rare Genetic Disorders and Multiscale

Contemporary reviews about precision medicine lay stress on the role of machine learning in rare genetic disorders identification and interpretation, focusing on the integration of learning features found in genomics, phenotypes, and clinical variables for improved yields in diagnoses [3]. Abbas et al. summarise the use of ML in improving variant prioritization, phenotype matching and risk stratification in rare disease contexts which prove beneficial in performance improvements using feature-rich, well-curated datasets [3]. Complementary methodological innovation in multiscale predictive modeling and digital twins suggest combining multi-omics and mechanistic hypothesis grammars to represent disease phenotypes and drug responses in silico to provide a route for taking molecular knowledge to personalised predictions for diagnostics and therapies [5]. Together, these contributions call for the inclusion of data-driven ML along with mechanistic models for the sake of translational impact.

AI Practices Bioinformatics and Healthcare Research Rigor

Broader examinations of AI in pharmaceutical sciences and clinical research focus on infrastructure, reproducibility and reporting best practices that are needed for clinical grade AI tools. Kandhare et al. review AI throughout the scope of pharmaceutical research, showing the way machine learning is speeding up the process of drug discovery and formulation, and advocating for standardized evaluations and validation protocols [7]. Sen and DeMazumder offer useful advice and a rigor check list for AI in healthcare research, which suggest methodological transparency, the use of proper baselines and reproducibility pipelines to ensure trustworthy outcomes in clinical studies [8]. These works in concert provide information that is crucial to the essential standards that would enable a safe transition of AI innovations from the research framework into clinical deployment: governance, reproducibility, and ethical standards.

Rule Based Inference, Explainability and Agentic AI in Clinical Systems

Hybrid diagnostic systems using rule-based inference with ML and explainable AI have become increasingly recommended in scenarios where systems for clinical settings have usability issues in terms of their transparency and cost. TS et al. present a dual-level dengue diagnosis using lightweight neural models with rule-based inference and explainable AI (XAI) in fog computing environments to display the hybrid systems ability to be efficient and interpretable [9]. Banerjie et al. Survey agentic artificial intelligence in healthcare: architectures, taxonomies, application domains, ethical and safety issues of autonomous decision-support agents in clinical practice [10]. Together these studies provide backing for design choices for the combination of rule driven logic with explainability and agentic safeguards for ethics reasons that address performance, interpretability, and clinical accountability.

III. PROPOSED METHODOLOGY

A. Data Acquisition Process and Preprocessing

The proposed system provides integrated structured data acquisition pipeline for genetic disorder prediction via machine learning driven inference. The dataset covers patient demographics, familial inheritance, severity level of the symptoms, biomedical measurements and factors under birth. These inputs are subjected to pre-processing such as dealing with missing values, numerical encoding and normalisation. A pre-trained scaler is used to scale all the features standardizing the model input and enhancing the prediction stability. Additionally, categorical attributes such as categories of symptoms and modes of inheritance are converted to numbers to be used in a model for compatibility. This preprocessing pipeline has ensured that noisy and data representations with various clinical heterogeneous data and data are converted into a structured representation known as 26 feature format for the classifier can still be reliable to predict with accurate value on different patient profiles.

B. Feature Engineering & Rule-Based Mapping

In order to improve the model interpretability and also be more relevant for diagnosis, a rule-based mapping is included in the feature engineering process of the system. Two engineered features, genetic disorder type and disorder subclass, are based on the category of symptoms and pattern of inheritance using domain knowledge. This mapping serves to cluster neurological, metabolic, blood-related and physical symptoms into meaningful diagnostic pathways that are in line with mitochondrial, multifactorial and single gene causing diseases. The hybrid method improves the predictiveness by introducing deterministic medical logic into an otherwise pure statistical model. Additional parameters such as autopsy, anomalies, IVF/ART and exposure indicator are given default clinical values: where user input is not applicable. This approach is effectively somewhat of an "input dimensionality breathing" approach that maintains the dimensionality of the input while retaining the meaningful clinical context present in the model decision process.

C. Developing Machine Learning Model

A Decision Tree classifier algorithm was chosen because of the interpretability and low computational complexity and for their suitability in healthcare decision support tasks. The model was trained from the structured dataset and hyperparameter tuning parameters were factored in in order to minimise overfitting and maximise generalisation. Feature importance analysis made sure that the important medical indicators (e.g., parental inheritance, blood test results, number of symptoms, and birth defects) were well captured. After the model was trained, it was serialized using pickle, in order to easily integrate the model in the backend of a web application developed with Flask. A separate scaler object was saved in order to ensure that the preprocessing is consistent in inference. The resulting model predicts one of nine genetic disorders in a high-accuracy classification with the ability to be deployed in a lightweight web-based environment in real-time.

D. Back End Processing / API Workflow

The frontend is implemented with the usage of the Flask framework which orchestrates the process of user authentication, session management, validation of user inputs, processing of predictions, and the communication with the generative AI module. Upon receipt of user input, the system generated the 26 feature vector, scaled the data, and sent them to the ML model. The output of the prediction is accompanied by a measure of confidence and disease label. Following this, an automated prompt is directed to Google Gemini to come up with medically contextualized recommendations. The back-end routes handle and secure the read/write operations to a user repository stored as a collection of user profiles in form of a JSON. The prediction API returns structured responses in the form of a json object, which makes it easy to make an API connection with a front-end interface and provides for efficient performance in real-time.

E.Generative/Clinical Recommendation Engine Powered by AI

To enhance the interpretability and usability by the patient, the machine combines Google's Gemini model as a medical guidance layer. After the operations in the ML model have predicted the class of disorder, a structured prompt is created that includes the patient's demographics, number of symptoms, indicators of inheritance and the disorder for which the patient had tested positive for. Gemini uses this information to give an actionable summary of the clinical information for the clinically relevant information such as explanation of the disorder, expectations for symptoms to expect, emergency indicators, lifestyle recommendations and genetic counseling recommendations. This generative module allows more trust, accessibility, and context to non-technical users by interpreting machine learning output into medically aligned narrative of insights. The hybrid combination of predictive analytics and natural-language reasoning makes the system a complete diagnostic aid tool rather than a classifier of such.

F. System Architecture

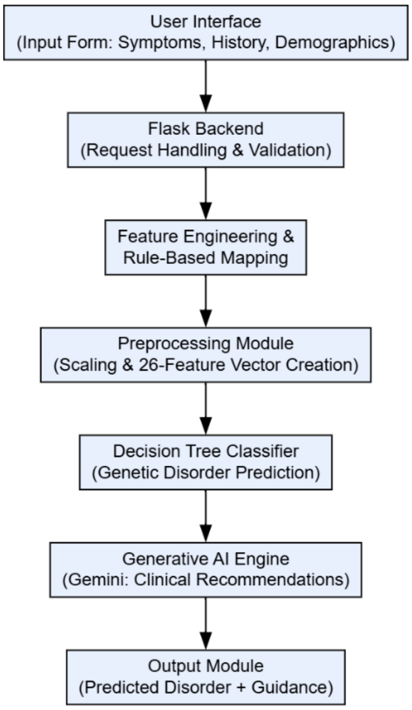


Fig 1 System Architecture

The system architecture is in the form of a modular workflow, which goes from the top down and includes user interaction, feature mapping, model inference, and generative recommendations layers. The process is triggered at the user interface where the inputs are got and are transferred to the back end (Flask) using secure routing end points. Preprocessing, scaler is performed and backend sends validated data to Decision Tree classifier for making a prediction. The label of the predicted disorder is then fed to the generative AI layer creating contextual recommendations to the user through a dynamic interface. Supporting modules: Including a user management, uses a JSON that is ensuring the reliability of the authentication and persistence of data. This multi-layered architecture is a cohesive architecture that allows for the genetic disorder screening to be done in real-time, in an interpreable, and user-planar mode.

IV. RESULT AND DISCUSSION

A. Registration Interface and User Onboarding User Interface

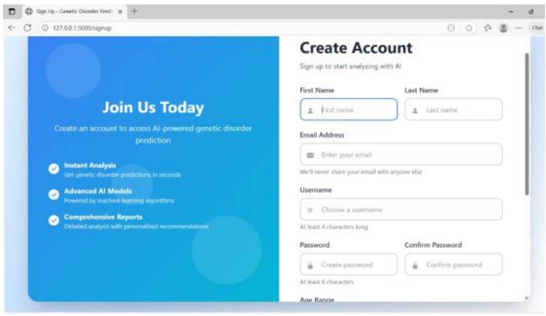


Fig 2: Registration Interface

The results shows that system is a clean, intuitive and user friendly registration interface with the aim of aiding accurate user onboarding. As shown in the first screen shot, the sign up page consists of necessary fields like Name, Mail id ,User name, Password, age range and the purpose of use. The layout involves the integration of defensable introducing panel with important features including instant analysis, advanced Jordi AI models and in-depth reporting. Real-time feedback mechanisms for example for password matches too increase usability and reduce input error The design is based on modern design principles for UI, and has been constructed with clarity and accessibility for use with different age groups. This onboarding experience is extremely important in terms of aiming to lower the barriers to the system's adoption and to ensure that the quality of the user-submitted data is high, in order to make accurate predictions.

B. Log- In Confirmation

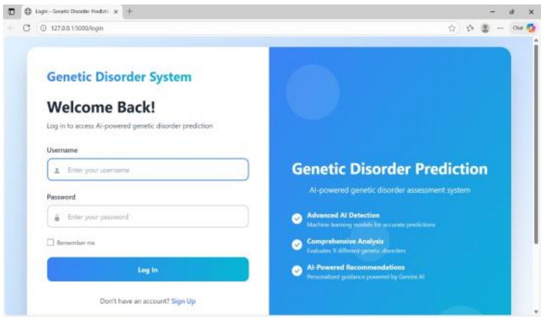


Fig 3: Login Confirmation

The second Screeenshots and depicts how to do a successful authentication and gaining seamless transition to the user dashboard. The interface provides personalized welcome message to the user, after log in and provide options for the user to navigate the system, eg, access to Patient Assessment Form and the log out button. The layout follows a constant visual styling, which is a part of a unity in a user experience. This stage is used to validate not only the security and correctness of the login mechanism, it's also used to validate if the system can still maintain user's session across the interactions. By only allowing access to authenticated access, the platform can make sure that sensitive prediction data is secured. The clean navigation interface will ensure that the users will be able to quickly find the assessment form without being confused, which will promises efficient workflow and uninterrupted use of the system.

C. Patient Assessment Form Interface/ Data Input Workflow

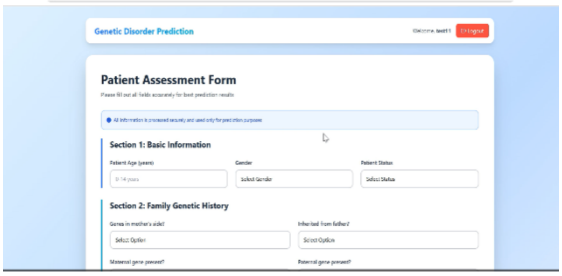
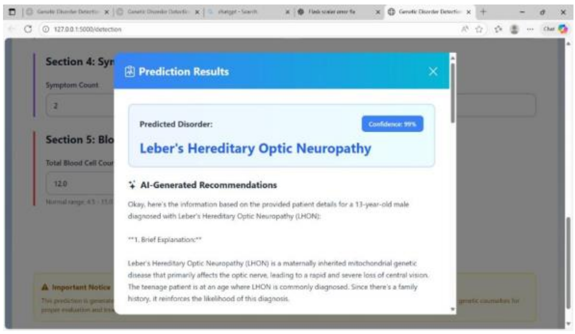


Fig 4 : Patient Assessment form

The third piece in the screenshots is the full Patient Assessment Form which is divided into several parts: Basic Information Family Genetic History Symptoms and Blood Test Indicators. Each section is nicely color coded and labeled for the sake of clarity. The interface reinforces secure data handling by an informational banner in order to assure users that all messages are handled in a confidential fashion. The form is designed to use dropdown menus for categorical attributes in order to reduce the ambiguity of input, and increase the consistency of the responses of different users of the questionnaire. The organization of the layout based on the long list of features used by the machine learning model, making sure that the user will enter full and proper data required for making the accurate prediction. This homework of the organized input capably reducing the user's effort supporting data reliability and model performance.

D. Prediction Cost and disorder detection Output



Prediction Results

Symptom Count: 2

Total Blood Cell Cour: 12.0

Predicted Disorder: Leber's Hereditary Optic Neuropathy

Confidence: 99%

AI-Generated Recommendations

Okay, here's the information based on the provided patient details for a 13-year-old male diagnosed with Leber's Hereditary Optic Neuropathy (LHON):

1. Brief Explanation:

Leber's Hereditary Optic Neuropathy (LHON) is a maternally inherited mitochondrial genetic disease that primarily affects the optic nerve, leading to a rapid and severe loss of central vision. The teenage patient is at an age where LHON is commonly diagnosed. Since there's a family history, it reinforces the likelihood of this diagnosis.

Fig 5: Detection of output

The fourth screenshot emphasizes the capabilities of the system to make real-time predictions based on the submitted inputs by users. In the example below, the model manages to identify "Leber's Hereditary Optic Neuropathy" with a confidence score of 99%. The outcome is shown in an output modal window where there is a good visual focus on the predicted disorder name, confidence indication, and recommendation process. The design of the responsive popup is to keep users from navigating away from the assessment form in order to view results of their assessment. The quality of the output can be seen with regard to the effectiveness of the machine learning classifier and how it matches the biomedical pattern mapping describing the patterns. This instant feedback mechanism builds high trust by the user and gives meaningful insights for further decision-making or medical consultation.

E. AI Clinical Explanation/Suggestions USCulation Print Out

In this it shows the extended AI generated recommendations which have come into being by the way of bringing together the Google's Gemini model. The system contains rich medical information that includes explanation of the disorder, symptoms to be monitored, lifestyle immigration, emergency signs and genetic counseling. The neat formatting and the organized formatting make sure that it is easily readable and the presence of an important disclaimer creates an underlined realization of the ethical boundaries of the system existed. The print preview ensures that the recommends can be exported into a formal document format which can be used for clinical documentation, research purposes or used to follow up discussions with the healthcare professionals. This capability is one of the testaments of the proof of strength of the system not only to predict, but also to bring the context-aware medical recommendation.

F. Discussion

The result of all the collected results give validation for the effectiveness of the proposed hybrid diagnostic framework which is shown to be high in usability, technically reliable and offer informative predictive output. The results of the interface resulted that system keeps the principles of User centered Design with the result that data navigating is smooth, data entering is precise and the authentication of user. The prediction result provided by the ML model has a good performance and can get a good disorder identification with the confidence scoring to initiate the user understanding and decision The combination of generative and AI allow the diagnostic experience to be taken to a higher level, the outputs of classifiers can be turned into rich and human-readable info, as the gap between what an AI is saying and how the doctor is interpreting it. In addition to this, having printable reports is something that brings to the practicality to the real world of the system. Overall the results show that the method is able to combining machine learning, rule-based mapping and generative AI in successful tru to make holistic and easy-to-use genetic disorder analysis platform.

V. CONCLUSION

This research proves to be the development of a hybrid research system, the genetic disorder prediction system based on artificial intelligence that proves to be effective in the integration machine learning, clinical mapping systems based on rules, and generative artificial intelligence-based medical guidance. The decision tree classifier with the structured preprocessing as well as feature engineering provides an accurate and interpretable prediction of disorders. The user-centered Web interface coupled with the secure authentication and the real time feedback adds not only ease, but also practicality of the use. The addition of explanations produced by Gemini takes unadorned predictions and produces clinically useful and meaningful decisions, which helps in medically multiply. Overall the system provides a reliable fast and easy to use diagnostics support tool which is playing a contributing factor in an early screening and better understanding of hereditary diseases.

VI. FUTURE ENHANCEMENT

Future improvements may come in the form of the incorporation of large and varied genomic data sets into the model, both to increase the accuracy of predictions and generalizability. Expanding the system to provide clinician feedback and recommendations in real time in various languages, as well as on the move, will open up the system. Additionally, the use of explainable AI techniques can help to increase transparency and clinical trust concerning complex prediction outcomes.

REFERENCES

[1] Changalidis, A., Barbitoff, Y., Nasykhova, Y., & Glotov, A. (2025). *A Systematic Review on the Generative AI Applications in Human Medical Genomics*. arXiv preprint arXiv:2508.20275.

[2] Ghosh, A. N. (2025). *Disease Prediction using Generative AI*. IJSAT-International Journal on Science and Technology, 16(2).

[3] Abbas, S. R., Abbas, Z., Zahir, A., & Lee, S. W. (2025). *Advancing genome-based precision medicine: A review on machine learning applications for rare genetic disorders*. Briefings in Bioinformatics, 26(4), bbaf329.

[4] Abugabah, A., Shukla, P. K., Shukla, P. K., & Pandey, A. (2025). *An intelligent healthcare system for rare disease diagnosis utilizing electronic health records based on a knowledge-guided multimodal transformer framework*. BioData Mining, 18(1), 70.

[5] Gopukumar, S. T., Dwivedi, D., Soni, T. K., Natarajan, P. G., Ramkanth, S., Mitra, A., & Das, U. (2025). *Multiscale Predictive Cellular Modeling: Integrating Hypothesis Grammars, Digital Twins, and Multi-Omics for In Silico Oncology and Precision Theranostics.*

[6] Alvi, R., Zaman, S. B., Karim, W., Abian, A. I., Raiaan, M. A. K., Mukta, S., ... & Azam, S. (2025). *Generative Artificial Intelligence in Bioinformatics: A Systematic Review of Models, Applications, and Methodological Advances*. arXiv preprint arXiv:2511.03354.

[7] Kandhare, P., Kurlekar, M., Deshpande, T., & Pawar, A. (2025). *Artificial Intelligence in Pharmaceutical Sciences: A Comprehensive Review*. Medicine in Novel Technology and Devices, 100375.

[8] Sen, C. K., & DeMazumder, D. (2025). *Getting Started on Artificial Intelligence in Health Care and Clinical Research: Includes Rigor Checklist for Authors and Reviewers*. Advances in Wound Care.

[9] TS, P. K. (2025). *Dual level dengue diagnosis using lightweight multilayer perceptron with XAI in fog computing environment and rule based inference*. Scientific Reports, 15(1), 1–27.

[10] Banerjie, S., Zhu, Y., Freeman, I., Machado, J. V., Ahmed, A., Sarker, A., & Al-Garadi, M. (2025). *Agentic AI in Healthcare: A Comprehensive Survey of Foundations, Taxonomy, and Applications*. Authorea Preprints.