RESEARCH ARTICLE

RSP Science Hub

International Research Journal on Advanced Science Hub
2582-4376
www.rspsciencehub.com
Vol. 07, Issue 06 June

Check for updates

http://dx.doi.org/10.47392/IRJASH.2025.062

A Comprehensive Risk Assessment of Genetical Disorders in Children

Dr. U Chaitanya[1], Shaik Nazrin Tarannum[2], Sushmitha Polishetty[3]

[1]Assistant Professor, Dept. of IT, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, India.

[2,3]UG Student, Dept. of IT, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, India.

Email ID: *uchaitanya_it@mgit.ac.in[1], nazrintarannumshaik@gmail.com[2], polishettysushmitha512@gmail.com[3]*

Article history

Received: 08 May 2025
Accepted: 20 May 2025
Published: 27 June 2025

Keywords:

Artificial Intelligence, Age Detection, Dynamic Questionnaire, Facial Age Detection, Genetical Disorder Risk Prediction, Machine Learning, Personalized Healthcare.

Abstract

The Genetical Disorder Risk Assessment System presents a web-based application of genetic disorder risk assessment through AI-powered face analysis and dynamic questionnaires. The system seeks to offer early, accessible, and personalized information regarding possible genetic health risks. Registration and face detection- based identification are performed using a webcam. The facial features detected are analyzed to estimate the age group of the user. Depending on the age detected, a suitable questionnaire is dynamically selected. The questionnaire gathers important information regarding family history, lifestyle, and general health indicators. An educated machine learning model processes the user's answer to estimate the risk of genetic disorders. The model provides a percentage risk value and labels the outcome as high, medium, or low. The system doesn't depend on uploaded medical records, increasing convenience and ease of use. Age based detection on faces guarantees that questionnaires are age relevant and tailored. The user-friendly web interface supports health awareness among non-expert users. Instant feedback enables users to take early preventive health measures. The project focuses on privacy ensuring safe processing and storage of user information. By integrating facial recognition, dynamic form creation, and AI based analysis the system offers an innovative method of early diagnosis. The tool has the ability to assist individuals and healthcare professionals in proactive genetic health.

1. Introduction

The field of genetic diagnostics has undergone a transformation through intersection with computer science and advanced sequencing technologies. Genetic disorders can manifest from abnormal or mutated DNA and can range from simple monogenic disorders to complicated polygenic disorders. Timely diagnosis is critical to successful management of these disorders, which will provide opportunity for early intervention, customized treatment strategies, and sometimes prevention. The classic approach to diagnoses—relating clinical assessment and family history—has experienced a significant transformation due to genomic technology offering accuracy and scalability [1]. Today's genetic diagnostics frequently use whole-genome sequencing (WGS), which allows the complete scrutiny of the genetic code of an individual. The sheer amount of data generated by WGS comes with challenges in interpretation, but these are being applied in

solution to machine learning (ML) and artificial intelligence (AI) algorithms. These data-driven models excel at pattern detection, disease classification, and estimating the odds of genetic disorders with great accuracy [2]. Sophisticated computational methods such as link mining and chain classifiers have been used to find multilevel relationships in genomic data and diagnose a variety of disorders simultaneously [3][7]. AI-especially decision trees, support vector machines (SVM), convolutional neural networks (CNN), and recurrent neural networks (RNN)- have outperformed traditional statistical models in predicting genetic disorders [4] [6]. There are important factors, such as automation, flexibility, and nonlinear high-dimensional datasets. Other comparative studies of different AI models of have shown high accuracy, sensitivity and specificity especially using gene expression and WGS [4] [5]. The utilization of ensemble learning and hybrid methods also indicates an increase in predictions especially for multifactorial disorders where genetic factors combine with environmental factors [5]. Genome sequencing has also significantly impacted pediatric medicine. Clinically, the use of WGS is increasingly employed to identify rare or undiagnosed conditions for infants and children, especially those who originally failed common diagnostic tests ([7]). This is of importance in cases of intellectual disability or general developmental delay, where complete assessments are essential for providing a complete diagnosis and effective treatment ([8]). AI is critical for interpreting genomic data, as it affords an automated approach that reduces the workload for clinicians and increases the speed of diagnosing patients ([9]). Despite these advances, prediction of genetic disorders continues to face challenges involving limited access to labeled datasets for rare conditions, that genetic data includes privacy issues, and the desire and need for explainable AI models that are able to corroborate medical decisions. Current studies will continue to wrestle with these issues in the hopes and of building better, more transparent, and more generalizable AI systems [10]. Recent developments in artificial intelligence (AI) and machine learning (ML) have revolutionized medical diagnostics, especially in the field of pre-symptomatic detection and risk identification in heritable and congenital disorders in pediatric populations. AI and ML allow novel predictions and personalized health opportunities by conceptualizing complex biological data. Literature states that AI methodologies are establishing increasingly robust support for the diagnosis of congenital conditions in pregnancies and for newborns regarding neonatal neurological conditions thereby showing integrity as a viable support for decision-making with regard to the clinical assessment process [11]. ML models and machine-learning classifiers have also been shown to evaluate the hereditary risk of genetic disorders in children which facilitate early prevention and assist parents in making informed clinical decisions [12]. Research evaluating the application of ML techniques for clinical support in predicting psychiatric and other hereditary disorders has been collected [13] showing the growth of ML present across a plethora of medical domains. In addition to heritable disorders, concurrent ML approaches have been used to predict chronic disorders like cardiovascular diseases indicating the opportunity presented through the augmentation of AI pathotyping [14]. Pairing big data analytics with ML will provide increased accuracy and applicability for large-scale disease prediction if data from around the globe are used to allow for real-time analysis and support of members of health communities [15].

2. Methodology

2.1. Genetical Disorder Risk Assessment System Architecture

The intended system is an intelligent, webpage-based system designed to determine the likelihood of genetic disorders in children and adolescents. This differs from other methods of obtaining a genetic risk/probability using large amounts of medical information or genetics testing. Our web-based system is non-invasive, and uses information based on the client/user and then converts that to a questionnaire based on the client's/user's age range using detected age from video captured through a webcam. The main idea is to captured a person’s face on a webcam to calculate an approximate age based on the analysis of that face. Once the approximate age is determined, the questionnaire will be generated and related to that age group. Upon completion of the questionnaire, the data the client/user supplied is run through a trained machine learning predictive model (Random Forest and Deep Neural Network), to determine the

probability/risk of having or developing a genetic disorder. The risk is categorized as high, medium, or low, and by percentage-based risk. This type of risk would allow clinicians to determine if a risk was detected, regardless of whether genetic testing was obtainable.

The components of the Genetical Disorder risk detection system are as follows:

- **Face Detection and Age Approximation-** This will be completed using OpenCV and pre-trained deep learning models that will use a webcam to determine the client's/user's age.
- **Dynamic and Responsive Questionnaire-** The system will generate a relevant list of questions based on the clients/users detected age group, and primarily focus on symptoms.
- **Risk Prediction-** Machine learning algorithms (Random Forest, DNN) trained on synthetic or available datasets to estimate the likelihood of a genetic disorder.
- **User-Friendly Interface-** A responsive web front-end evolved to provide a seamless user experience and accessibility.
- **Secure User Management-** Registration and login features including optional child and parent user modes for different age ranges.

This system does not require the upload of medical or genetic reports, protecting privacy, and providing barrier-free access. Ultimately, the aim is to provide users to access; particularly parents wanting earlier indications of possible genetic disorders in their children; and to seek timely medical intervention if needed. (Figure 1) The system architecture diagram outlines the steps of a web application that performs risk assessment of genetic disorders risk detection system. The initial steps include providing either a user registration or logging in, followed by a webcam based facial capture for the systematic identification and estimation of age through a machine learning trade space exploration face detection and age detection model. After the age prediction is performed, for users of different age groups, the system dynamically provides a questionnaire with specific pre-selected questions based on their predicted age group. The user will answer the questions from the questionnaire, and their responses will go through a risk prediction process which predicts the likelihood of them being at risk for a genetic disorder, and classifies the risk as low, medium, or High

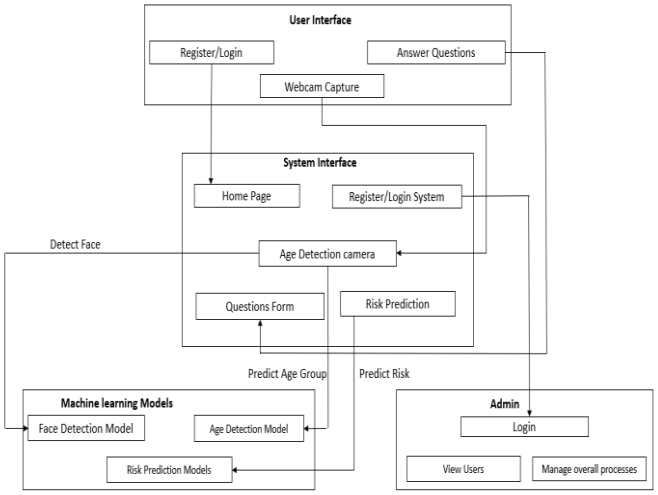

Figure 1 System Architecture of Genetical Disorder Risk Assessment System

2.2. Dataset

Rather than gathering real time medical or genetic information from medical facilities, we created a synthetic dataset based on a computerized replication of survey responses. The data was created using medically validated symptoms, risk factors, and family history indicators for common genetic disorders in children. This method of data creation allowed us to rapidly simulate many different realistic user profiles across the several different target ages groups. We created samples to train and test the machine learning models. The completions are broken into three age-based categories: infants (0–2), children (3–12), and adolescents (13–18). Every sample is created with the responses to dynamically generated questions that will pertain to genetic predisposition, developmental indicators lifestyle, and environmental exposures.

2.3. Data Pre-processing

Data preprocessing is an important step which is necessary to prepare the synthetic questionnaire dataset before training and testing the machine learning models. The raw dataset consists of simulated responses provided by users across various age groups, and after some data pre - processing we were able to get the data in a state that allowed us to evaluate the quality of the data and the reliability of the dataset as input for our models. The

various preprocessing tasks included the following: the categorical variables of family history, consanguinity, and developmental delays, were one-hot encoded to account for categorical data rather than ordinal data input for any of the machine learning models. The numerical variables of parental age at conception, and risk percentage were normalized using min-max scalers to ensure that these particular variables were kept within a standard numerical range to encourage convergence and to improve performance, across the input. (Figure 2)

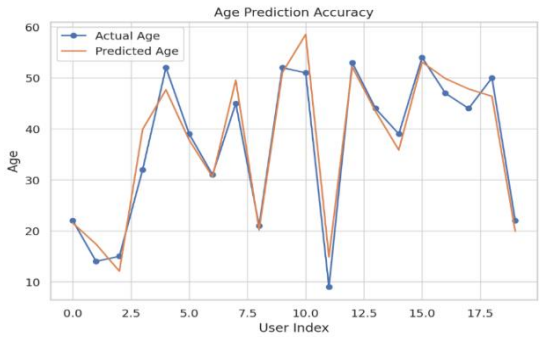

Figure 2 Age Prediction Accuracy of Genetical Disorder risk assessment System

The Figure 2 above illustrates the accuracy of predicted age against the actual age of users, indexed from 0 to 18. The blue line with circular markers represents the actual ages, while the orange line shows the corresponding predicted ages generated by the facial age estimation model

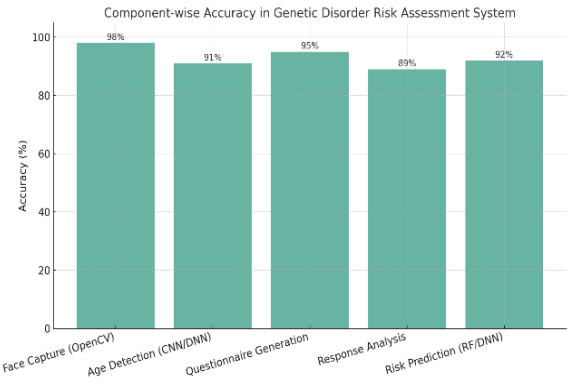

Figure 3 Accuracy of Each Component in Genetical Disorder Risk Assessment System

The Figure 3 illustrates bar chart shows the precision of every core component in the genetic disorder risk assessment system. Face capture with OpenCV attains the highest level of accuracy at 98% providing table image input.

3. Results and Discussions

The facial age prediction model was tested on a set of 20 users. The actual ages and the predicted ages were graphed to determine the performance of the model. As seen from the graph, the predicted age values trace the actual age values closely with little deviation for the majority of user instances. The model was able to predict age correctly within a ± 5year range for most users. (Figure 4)

Figure 4 Home Page

The Figure 4 provides a welcoming introduction to the Genetic Disorder Risk Assessment System. It guides users to register or log in, explaining the purpose and benefits of the application.

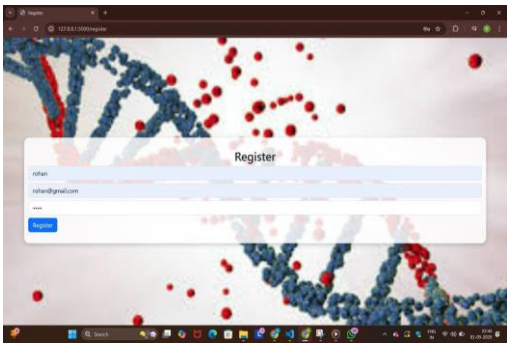

Figure 5 Users Registration

The Figure 5 User Registration page allows new users to create an account by entering basic details such as name, email, password

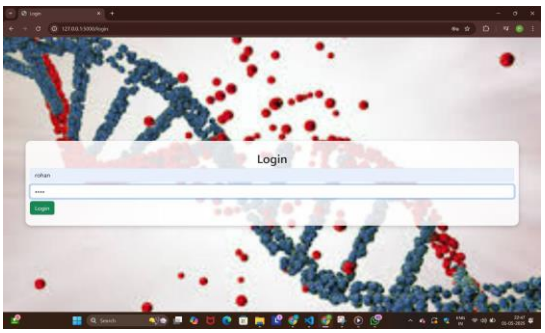

Figure 6 User's Login

The Figure 6 User Login page enables registered users to securely access their accounts using their email and password.

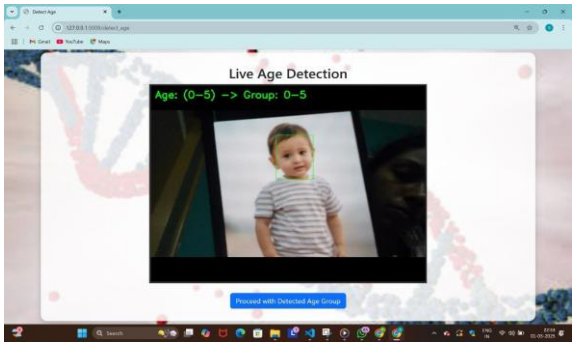

Figure 7 Age Detection Through Face

The Figure 7 page displays the detected age of the user, identified through facial recognition, and confirms that the user falls within the 0–5 age range.

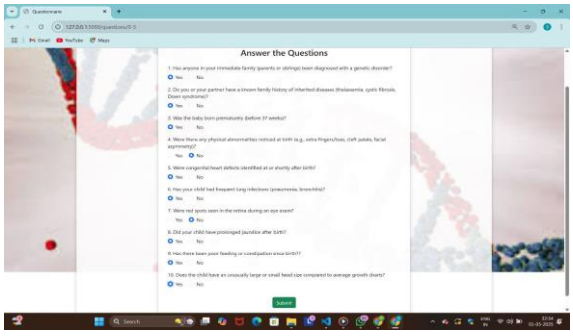

Figure 8 Questionnaire Generated Based on the Age Detected

The Figure 8 presents a dynamic questionnaire tailored to the user's detected age group. The questions are relevant to the user's developmental stage and are designed to assess potential genetic disorder risks effectively.

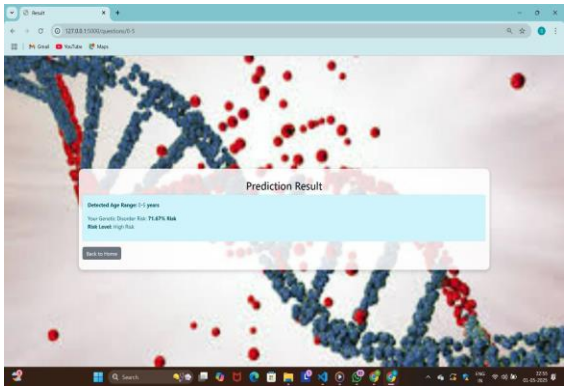

Figure 9 Displaying the Result of Genetic Risk Present in the Body

The Figure 9 displays the final risk assessment results based on the user's questionnaire responses. It clearly shows the risk level High, Medium, or Low along with the calculated percentage.

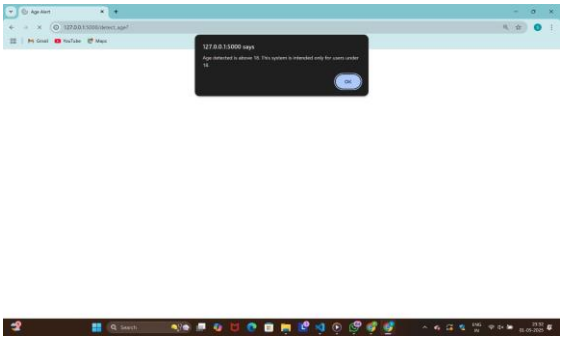

Figure 10 Popup Displayed for the User Whose Age is Out of Range

The Figure 10 displays the popup message which was given for the user whose age is out of range i.e., more than 18

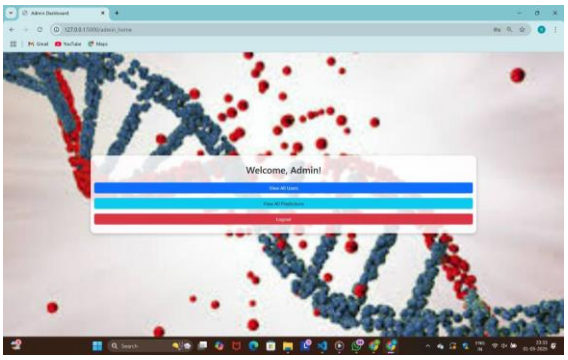

Figure 11 Admin's Dashboard

The Figure 11 Displays the Admin's dashboard with multiple options within it.

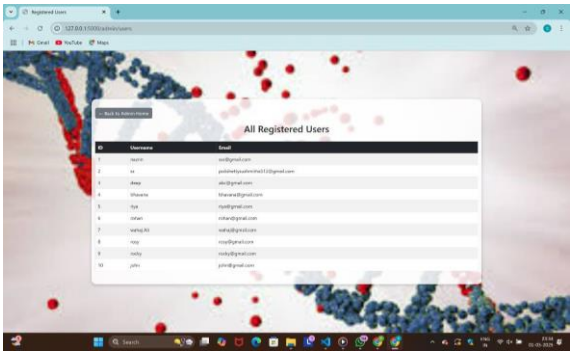

Figure 12 All Registered user's

The Figure 12 displays the registered users in the application which can only be seen by the admin.

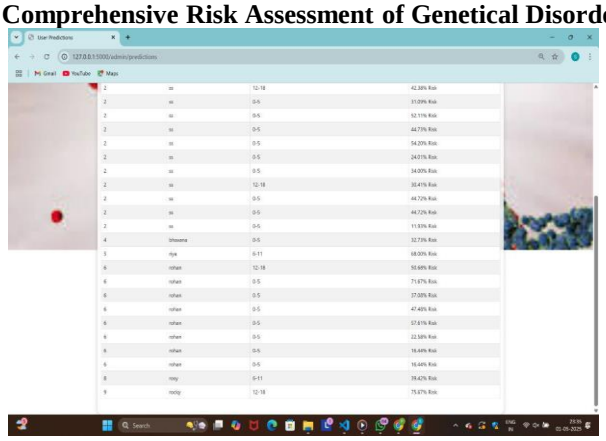

Figure 13 Page Displaying the Predictions made by the users

The Figure 13 displays the predictions made for different users who registered for our application which can be seen only by the admin itself.

Table 1 Risk Prediction Results Across Various Age Ranges

Test Cases	Actual Age	Predicted Age Range	Risk Percentage	Risk Level
01	5 yrs	(0-5)	82%	High
02	11 yrs	(6-12)	65%	Medium
03	13 yrs	(13-18)	43%	Low
04	22 yrs	(19-25)	Age out of range cannot be predicted.	Cannot be predicted.

The Table 01 above summarizes test cases from the genetic disorder risk assessment system, showcasing how the model predicts risk levels across different age groups. Each test case represents a user who underwent facial age detection followed by a dynamic questionnaire. Based on their responses, the system calculated a risk percentage and categorized it as Low, Medium, or High.

Conclusion

The web - based system designed for the risk assessment of genetic disorders provides a new and convenient solution combining AI-based technologies with ease of use. Through the implementation of OpenCV for real-time face detection, the system accurately captures user images and estimates age without requiring input, thus facilitating age related questionnaires, providing personalized and relevant data collection for effective risk prediction. By employing machine learning models like Random Forest and Deep Neural Networks constructed with Keras & Tensor Flow, the system efficiently processes user responses to determine the probability of genetic disorders. The models exhibit robust performance metrics, with high accuracy and reliability throughout all phases of the pipeline—ranging from face capture to ultimate risk classification.

References

[1]. Dwivedi, S. D., Yadav, S. D., Sahu, D., Singh, D., Rawat Singh, M. (2024). Revolutionizing genetic diagnostics: Innovative techniques for inherited disease detection. Gene Reports, 36, 101963. doi: https:// doi.org/ 10.1016 /j.genrep. 2024.101963.

[2]. Alzahrani, A. A., & Alharithi, F. S. (2024). Machine learning approaches for advanced detection of rare genetic disorders in whole-genome sequencing. Alexandria Engineering Journal, 106, 582-593. doi: 10.1016/j.aej.2024.08.056.

[3]. Sen, M., Rajasekaran, R., JayaRam Reddy, A., & Govinda, K. (2024). Predicting Genetic Disorders: A Link Mining Approach. 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE),15. doi:10.1109/ IITCEE59897. 2024.10467830.

[4]. Chaplot, N., Pandey, D., Kumar, Y., Sisodia, P. S. (2023). A comprehensive analysis of artificial intelligence techniques for the prediction and prognosis of genetic disorders using various gene disorders. Archives of Computational Methods in Engineering, 30(6), 3301–3323. doi: 10.1007/s11831-023-09904-1.

[5]. Solomon, D. D., Sonia, K., Kumar, K., Kanwar, K., Iyer, S., Kumar, M. (2024). Extensive review on the role of machine learning for multifactorial genetic disorders prediction. Archives of Computational Methods in Engineering, 31(2), 623–640. doi: 10.1007/s11831-023-09996-9.

[6]. [6] Atta-Ur-Rahman, Nasir, M. U., Gollapalli, M., Zubair, M., Saleem, M. A., Mehmood, S., Khan, M. A., Mosavi, A. (2022). Advance genome disorder prediction model empowered with deep learning. IEEE Access, 10,70317–70328. doi: 10.1109/ACCESS.2022.3186998.

[7]. Raza, A., Rustam, F., Siddiqui, H. U. R., Diez, I. D. L. T., Garcia-Zapirain, B., Lee, E., Ashraf, I. (2023). Predicting genetic disorder and types of disorder using chain classifier approach. Genes, 14(1), 71. doi: 10.3390/genes14010071.

[8]. Bick, D., Jones, M., Taylor, S. L., Taft, R. J., Belmont, J. (2019). Case for genome sequencing in infants and children with rare, undiagnosed or genetic diseases. Journal of Medical Genetics, 56(12), 783-791. doi: 10.1136/jmedgenet2019-106111.

[9]. Moeschler, J. B., Shevell, M., Committee on Genetics. (2014). Comprehensive evaluation of the child with intellectual disability or global developmental delays. Pediatrics, 134(3), e903-e918. doi: 10.1542/peds.2014-1839.

[10]. Dias, R., Torkamani, A. (2019). Artificial intelligence in clinical and genomic diagnostics. Genome Medicine, 11(70). doi: 10.1186/s13073-019-0689-8.

[11]. Kaur, K., Singh, C., Kumar, Y. (2023). Diagnosis and detection of congenital diseases in newborns or fetuses using artificial intelligence techniques: A systematic review. Archives of Computational Methods in Engineering 30, 3031–3058. doi: 10.1007/s11831-023-09892-2.

[12]. Ramaswamy, M., Senthilnathan, S., Rajan, A., Saravanan, I., Snehal, M., Siva,K. (2024). Machine learning model to predict risk assessment of a child inheriting a genetic disorder. bioRxiv. doi: 10.1101/2024.08.22.609095.9

[13]. Bracher-Smith, M., Crawford, K., Escott-Price, V. (2021). Machine learning for genetic prediction of psychiatric disorders: A systematic review. Molecular Psychiatry, 26, 70– 79. doi: 10.1038/s41380-020-0825-2.

[14]. Mohan, S., Thirumalai, C., Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542– 81554.doi: 10.1109/ACCESS.2019.2923707.

[15]. Chen, M., Hao, Y., Hwang, K., Wang, L., Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. IEEE Access, 5,8869–8879. doi: 10.1109/ ACCESS. 2017. 2694446.