ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.206 **Volume 8, Issue 7, July 2025**

© 2025 IJMRSET | Volume 8, Issue 7, July 2025| DOI:10.15680/IJMRSET.2025.0807054

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Machine Learning- Based Analysis of Genetic Chaos Theory

Shalini G[1], Evanjalin Issabella. Y[2], Praisy Jasmine P[3], Harini L[4]

Department of Computer Science and Engineering, Karpaga Vinayaga College of Engineering and Technology, Chengalpattu, Tamil Nadu, India[1-4]

ABSTRACT: Machine learning is an application of artificial intelligence that provides systems that ability to automatically learn and improve experience without being explicitly programmed. We suggest a methodology for predicting biological anomalies in this work. We focused on estimating inherited abnormalities because we are aware that each of our biological parents contributes half of our inherited traits. As a result, we currently have a wealth of information about genetic illnesses, and technological advancements are accelerating. Inefficiently, our proposed method assesses chromosomal abnormalities via a classification algorithm in supervised technique. Patterns of behavior, neurological abnormalities, visual or hear loss, stunted growth, chronic diseases, birth defects and others can all be explored. Our suggested strategy can be more effective because it aids in the discovery of genetic abnormalities while requiring much less time. A Random Forest Algorithm was required to predict which autoimmune mutations are likely to appear in the coming years in order to predict a positive experience with our model. This model which predicts an individual's of developing all known genetic mutations have been developed by analysts.

KEYWORDS-machine learning, biological anamolies, supervised learning, random forest

I. INTRODUCTION

The Project aims to implement a robust Machine learning model that can efficiently predict the disease of human, based on the symptoms that he/she possesses. Machine learning algorithms are performing very well, the approaches will helps us to keep the predictions much more accurate on completely with complex data. So what is machine learning and how can it provide accurate results in efficient manner, Machine learning is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The primary aim is to allow the computer to learn automatically without human intervention or assistance and adjust actions accordingly. The test for a machine learning model is a validation error on new data not a theoretical test that proves a null hypothesis. Because this often uses an iterative approach to learn from data, the learning can be easily automated.

II. MACHINE LEARNING METHODS

Some of the methods of Machine Learning algorithm

a) SUPERVISED LEARNING

A Supervised learning algorithm learns from labeled training data, helps you to predict outcomes for unforeseen data. It is highly accurate and trustworthy method.

b) UNSUPERVISED LEARNING

Unsupervised learning algorithm is the type of self - organized with the help of previously unknown patterns in dataset without pre-existing labels.

c)SEMI-SUPERVISED LEARNING

Semi- supervised learning is the combination of both supervised and unsupervised which mean labeled and unlabeled data.

© 2025 IJMRSET | Volume 8, Issue 7, July 2025| DOI:10.15680/IJMRSET.2025.0807054

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

d) REINFORCEMENT MACHINE LEARNING Reinforcement machine learning is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.

PURPOSE OF THE PROJECT

Our model helps primary care properly identify biological anomalies in the biomedical field, where machine learning already has a substantial impact in a number of fields. Let's say there is a person with respiratory or neurological issues, and we are unsure if that individual has a neurological mutation. Now that we have this model, we can use it to anticipate events with the best degree of accuracy. It is difficult to assess and determine the precision of a genetic characteristic. Based on birth abnormalities, maternal genes, paternal genes inherited from the father, maternal genes, maternal genes on the mother's side, and blood test results, we utilize this model to forecast genetic illnesses. Both numerical and category target variables can be used with this approach. In this research, we provide an algorithm that learns from previous data and makes predictions. In order to successfully recruit the data set, this model aims to select the best fit model. Then, we forecast future data using the previous data with the highest accuracy. More importantly, our model could produce accurate forecasts and wise conclusions when compared to traditional methods.

SCOPE OF THE SYSTEM

It works in nonlinear problem and can be performed the structure and unstructured data. It is very commonly used in various industries such as banking, healthcare and media. A random forest generates accurate predictions that are simple to understand. It is adept at effectively processing huge datasets. The random forest algorithm significantly outperformed the decision tree method in the accuracy of predictions..

III. METHODOLOGY

1.Data Collection: Gather datasets containing genetic information, including chromosomal abnormalities, patterns of behavior, neurological abnormalities, visual or hearing loss, stunted growth, chronic diseases, and birth defects.
2.Data Preprocessing: Clean and preprocess the data to ensure quality and consistency.
3.Feature Selection: Identify relevant features that contribute to genetic abnormalities.
4.Random Forest Algorithm: Implement a Random Forest classification algorithm to predict the likelihood of autoimmune mutations.
5.Model Evaluation: Evaluate the performance of the model using metrics such as accuracy, precision, recall, and F1-score.

PROPOSED WORK

1.Feature Extraction: Extract relevant features from the data.
2.Model Training: Train the Random Forest model using the extracted features.
3.Prediction: Use the trained model to predict the likelihood of genetic abnormalities.
4.Output: Output the predicted results, highlighting potential genetic mutations.

EXPECTED OUTCOMES

1.Improved Accuracy: The proposed methodology is expected to improve the accuracy of predicting genetic abnormalities.
2.Reduced Time: The model is designed to require less time for predicting genetic abnormalities.
3.Effective Decision Support: The system will provide effective decision healthcare

FUTURE WORK

1.Model Optimization: Optimize the Random Forest model for better performance.
2.Integration with Clinical Decision Support Systems: Integrate the proposed methodology with clinical decision support systems for real-world applications.
3.Expansion to Other Genetic Disorders: Expand the methodology to predict other genetic disorders and abnormalities.

SYSTEM ANALYSIS INTRODUCTION

A regression is made for the usefulness of data collection, data analysis and data integration. The larger areas such as industries that are generating tons of data will get the help of regression to every element of data operations like data labeling, segmenting and analyzing. The fusion of regression with massive data is a neverending loop is analyzing the

© 2025 IJMRSET | Volume 8, Issue 7, July 2025| DOI:10.15680/IJMRSET.2025.0807054

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

best engine is the most critical element for a profitable business. Regression analyzes the best designs and furnishes the best since the engine is considered of the car. The massive data allows industries to calculate the probability of different outcomes and decisions. Predictive analysis helps then by providing a recommendation for manufacturing. The input of regression is the information extracted for massive data. Hereby making a regression of finding the best design among the miscellaneous designs, makes the industry advances in their business by providing a good quality product.

The System Development Lifecycle framework is designed to outline a complete development and implementation process suitable for developing complex applications. SDLC is a process followed for a software project, within a software organization. It consists of a detailed plan describing how to develop, maintain, replace and alter or enhance specific software. The life cycle defines a methodology for improving the quality of software and the overall. The major inputs and outputs and major functions of the system are follows: Input: ∟ The vendor must create the account for login. All the vendor details and their company and engine design have been stored the datail in our database.In regression, the best fit line is the line that perfectly suited the given scatter plot. We started with a scatter plot to check the relationship between the dependent and independent variable using a straight line. We then reduce the RSS to find the equation of the best fit line and the optimal values of intercept and slope. We'll need a lot of time to figure out intercept and slope. Because we find out the m and c in a variety of ways, we can become confused when determining the slope and intercept. We also check the correlation or not .

PROBLEMS IN EXISTING SYSTEM

To deal with data, they proposed an evolutionary framework, an efficient algorithm called chromatin programming, and a least-squares estimator in the existing model. The model is fed propaganda such as neurological issues, depressive symptoms, breathing issues, and the prediction of a neurological mutation. We will first read and draw the scatter plot before preparing your data for linear modeling. Then, using the least squares method, determine the model's accuracy. The existing method checks for a correlation and then finds the equation of the most suitable line while diminishing the RSS and determining the optimal values. The solution necessitates a long time because this model is designed in a straight line to check the relationship between a dependent and independent variable. Every time we perform regression, We must determine whether or not the fitted line is significant. Then we look for residuals and ways to reduce the cost function. This model uses differentiation and gradient descent to find the slope and intercept, and then we find the fit line, so predicting the variable necessitates a lengthy process.

IV. PROPOSED SYSTEM

Even though we are aware that each of our biological parents contributes half of our chromosomes to our makeup, we concentrated on hereditary disease prognostication in this model, because it is challenging to accurately quantify and assess a hereditary trait. This model is developed to forecast chromosomal abnormalities based on blood test results, birth defects, maternal genes, paternal genes inherited from the father, and genes on the mother's side. A machine learning method has emerged as a potent computational tool for epigenetic diseases. We can definitely make do without knowledge and extrapolate data as well. Both categorical and numerical target variables can be used with this strategy.

ADVANTAGE OF PROPOSED SYSTEM

In this classification, it will help we make predictions in cases where the output is a categorical variable.

SYSTEM REQUIREMENT
Hardware requirements
Disk: 120 GB Monitor: 15’’ LED SVGA Input
Devices: Keyboard, Mouse
Software Requirements
Front end : Core Python, CSS, JS Web application : Django, Flask Back end : MySQL
Operating system
Coding Language: PYTHON IDE: PyCharm
Database: MySQL

© 2025 IJMRSET | Volume 8, Issue 7, July 2025| DOI:10.15680/IJMRSET.2025.0807054

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

INTRODUCTION TO PYCHARM

PyCharm is a popular integrated development environment (IDE) for Python programming. It provides a comprehensive set of tools and features to help developers write, debug, and test their code efficiently

KEY FEATURES OF PYCHARM

Code Editor: PyCharm's code editor provides syntax highlighting, code completion, and inspections to help developers write clean and efficient code.

Debugging: PyCharm's built-in debugger allows developers to step through their code, set breakpoints, and inspect variables.

Project Management: PyCharm provides a project management system that helps developers organize their code, manage dependencies, and track changes.

Version Control: PyCharm supports version control systems like Git, SVN, and Mercurial, making it easy to collaborate with others and manage code changes.

Code Analysis: PyCharm's code analysis features help developers identify potential issues, such as syntax errors, type errors, and security vulnerabilities

USER AUTHORIZATION

One way to perform session tracking is to leverage the information that comes with User authorization.

When a web server restricts access to some of its resources to only those clients that log in using a recognized username and password. After the client logs in, the username is available to a servlet through getRemoteUser (). When use the username to track the session. Once a user has logged in, the browser remembers her user name and resends the name and password as the user views new pages on the site. A servlet can identify the user through her username and they're by Track her session.

V. MODULES

1)Patient
2)Laboratory
3)Doctor
4)Estimator
5)Admin

Patient: This module gives the registration process with the patient details of name, Email id, phone no, passwords, address, and date of birth. With this, the patient can log in to the patient page. If the patient is new, then the patient creates a new account. After the login process, the patient fills in the name, address, disease, Email id, Date of birth, and Address on the patient page

Laboratory: This module fills in all the necessary information for registering with the lab. The user has to fill in the details of name, email id, password, and phone no. After which, they can log in to the lab's page. If the lab is a new one and does not have an account yet, then you need to create one as well on the lab's page

Doctor: This module offers a registration process for the doctor's details of name, email id, password, and date of birth. Along with this address and phone number can be given to log in to the doctor page. Once the admin approves providing the doctor access then the doctor can log in to their page.

Estimator: This module gives the registration process with the estimator details of name, email id, password, and phone no with the estimator can log in to the estimator page. If admin approves the estimator then the estimator login to the estimator page.

Admin: This model allows the admin to log in to the admin page. The patient fills in the patient details then the admin approves the patient details and sends them to the doctor. If the doctor views the patient report, then the admin approves the doctor.

© 2025 IJMRSET | Volume 8, Issue 7, July 2025| DOI:10.15680/IJMRSET.2025.0807054

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

SYSTEN TESTING AND ENVIRONMENT INTRODUCTION

Software testing is a critical element of software quality assurance and represents the ultimate review of specification, design and coding. In fact, testing is the one step in the software engineering process that could be viewed as destructive rather than constructive.

EXPECTED RESULT

1.Username is correct. Password is incorrect. Username and Password is incorrect. Username and Password is incorrect.
2.Username is incorrect. Password is correct. Username and Password is incorrect. Username and Password is incorrect.
3.Username is empty. Password is correct. Username is required. Username is required.
4.Username is correct. Password is empty. Password is required. Password is required
5.Both Username and Password is incorrect. Username and Password is incorrect. Username and Password is incorrect.
6.Both Username and Password is empty. Username and Password is required. Username and Password is required.
7.Both Username and Password is correct. Login Successful. Login Successful.

USER TESTING

Unit testing focuses verification effort on the smallest unit of software design.
Basic path testing:
1.The established technique of flow graph with Cyclamate complexity was used to derive test cases for all the functions
Loop Testing:
2.In this type of testing all the loops are tested to all the limits possible.

VI. CONCLUSION

The precision of our suggested model increases for chromosomal abnormalities. The decision's knowledge extraction and, consequently, decisions, will be more accurate the more data that are accessible for lifting the decision. Decision tree algorithms, which are used to make choices and are straightforward to use data mining techniques with high predictive accuracy, are one potential approach to abstracting knowledge from historical data. To affect our model, we suggest a uniform catalytic with chromosome symptoms and a random forest.

Which data have the highest accuracy. That is determined by the proposed classification technique. Our algorithm quickly eliminates unnecessary information while enhancing the findings' readability and precision. We are also excellent senior communication tools, and this model consistently forecasts sound judgement.

VII. FUTURE ENHANCEMENT

Our proposed model improves accuracy in chromosomal abnormalities. The more data available for lifting the decision, the more accurate its knowledge extraction, and thus its decisions, will be. A possible idea is to abstract knowledge from historic information into decision tree algorithms, which are used to make decisions and are simple to use data mining techniques with high predictive accuracy. We propose a Chromosome symptom homogeneous catalytic with a random forest to influence our model.

REFERENCES

1.S.-M. Udrescu and M. Tegmark, “Ai feynman: A physics-inspired method for symbolic regression,” Science Advances, vol. 6, no. 16, 2023.
2. J. Hung Yon, L. Muchen, and Y. Ong, “Gene expression programming: A survey,” IEEE Computational Intelligence Magazine, vol. 12, pp. 54– 72, Aug 2024.
3. P. G. Espejo, S. Ventura, and F. Herrera, “A survey on the application of genetic programming to classification,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 2,pp. 121–144, March 2023.
4.C. Ferreira, “Gene expression programming: a new adaptive algorithm for solving problems,” Complex Systems, vol. 13, no. 2, pp. 87–129, 2022

5. Y. Suttasupa, S. Rungraungsilp, S. Pinyopan, P. Wungchusunti, and P. Chongstitvatana, "A comparative study of linear encoding in genetic programming," in 2011 Ninth International Conference on ICT and Knowledge Engineering, Jan 2023, pp. 13–17

6. N. R. Sabar, M. Ayob, G. Kendall, and R. Qu, "Automatic design of a hyper-heuristic framework with gene expression programming for combinatorial optimization problems," IEEE Transactions on Evolutionary Computation, vol. 19, no. 3, pp. 309–325, June 2020.

7. Chi Zhou, Weimin Xiao, T. M. Tirpak, and P. C. Nelson, "Evolving accurate and compact classification rules with gene expression programming," IEEE Transactions on Evolutionary Computation, vol. 7, no. 6, pp. 519–531, Dec 2022

8. V. K. Karakasis and A. Stafylopatis, "Efficient evolution of accurate classification rules using a combination of gene expression programming and clonal selection," IEEE Transactions on Evolutionary Computation, vol. 12, no. 6, pp. 662–678, Dec 2019.

9. J. Zhong, L. Feng, W. Cai, and Y. S. Ong, "Multifactorial genetic programming for symbolic regression problems," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 50, no. 11, pp. 4492–4505, Nov 2021.

10. Z. Huang, M. Li, A. Mousavi, M. Danishva, and Z. Wang, "Egep: An event tracker enhanced gene expression programming for data driven system engineering problems," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 3, no. 2, pp. 117–126, April 2020

11. J. C. Sprott, Chaos and Time-Series Analysis, 2nd ed. Oxford: Oxford University Press, 2022

12. P. Cvitanović et al., Chaos: Classical and Quantum, 4th ed. Princeton, NJ: Princeton University Press, 2023

13. S. Banerjee and G. C. Verghese, Nonlinear Phenomena in Dynamical Systems: Chaos, Fractals, and Complexity, 3rd ed. Cham: Springer, 2023.

14. P. Baldi and S. Brunak, Bioinformatics: The Machine Learning Approach, 3rd ed. Cambridge, MA: MIT Press, 2024

15. L. da F. Costa et al., Complex Networks: Principles, Methods and Applications, 2nd ed. Cham: Springer, 2022.

16. Y. Sun, A. Martinez, and J. Z. Wang, AI for Bioinformatics: Advances and Applications. Cham: Springer, 2023.

17. M. Bohanec et al., Explainable Artificial Intelligence: Foundations, Methodologies, and Applications. Cham: Springer, 2023.18. R. L.

Haupt and S. E. Haupt, Practical Genetic Algorithms, 4th ed. Hoboken, NJ: Wiley, 2022.

19. S. Raschka, Y. Liu, and V. Mirjalili, Machine Learning with PyTorch and Scikit-Learn: Develop Machine Learning and Deep Learning Models with Python, 2nd ed. Birmingham, UK: Packt Publishing, 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com