



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IX **Month of publication:** September 2026

DOI: <https://doi.org/10.22214/ijraset.2026.84856>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Predicting the Structural Impact of Genetic Variants on 3D Genome Architecture Chromatin Interaction

Dr. T. Rajesh¹, Divya Sree², Sai Pranathi³, Neha Rao⁴

¹Associate Professor, CSE, GNITS, Hyderabad, India

^{2,3,4}Student, CSE, GNITS, Hyderabad, India

Abstract: Mutations and single nucleotide polymorphisms are examples of genetic variations which play a critical role in the development of the structure of the genome, which is critical for its regulation. Unlike having a linear structure, the genome is structured into three-dimensional structures where various regions of the genome, for example, enhancers and promoters can interact with each other. Chromatin interactions are crucial for gene expression patterns, and genetic variations leading to disruption of these interactions can cause chromatin folding and loops to become distorted and gene expression pattern to become irregular, thus causing illnesses such as cancer and genetic disorders.

Despite its importance for diagnostics, studying the effects of genetic variations on the structure of the genome poses a challenge in light of the genome's hierarchical and dynamic nature. The Hi-C approach generates huge volumes of chromatin interaction data, which cannot be easily analyzed using conventional computational methods due to high dimensionality. Additionally, conventional computational algorithms lack efficiency in analysis of large data volumes as well as ability to find nonlinear relationships among different genetic elements.

To address these challenges, this study proposes a machine learning approach to estimate structural disruptions resulting from genetic mutations based on the information of chromatin interactions. The genomic features used in the model are inter-action frequencies, genomic distances between two binding sites, and relative locations of a mutation in relation to nearby regulatory regions. Machine learning algorithms like Support Vector Machine and Random Forest capture patterns of structurally disruptive mutations and enable predictions regarding the effect of mutations on structures.

This study finds that the proposed approach is capable of identifying structurally disruptive mutations and providing insights into the structure of the genome and biological processes relevant to the disease. This approach enables scientists to prioritize genetic variations based on their potential to disrupt the structure of the genome. Thus, this approach can be viewed as a viable substitute for traditional approaches and advances genomics and precision medicine.

Index Terms: Genetic variants, single nucleotide polymorphisms (SNPs), 3D genome organization, chromatin interactions, Hi-C data, machine learning, feature extraction, structural variation, gene regulation, precision medicine.

I. INTRODUCTION

The genome is three-dimensional by nature. The three-dimensional architecture allows distant parts of the genome to contact each other, an essential aspect of gene regulation via enhancer-promoter communication. Aberrant folding or looping of DNA molecules leads to disruption of normal gene regulation. Thus, genome interactions play a crucial role in transcriptional activation and proper cellular function. Genome variations such as SNPs or mutations can disrupt genome interactions and alter its architecture. Changes in chromatin architecture may result in abnormal gene expression and are frequently associated with pathologies, including cancer and hereditary conditions. Thus, understanding the effects of genome variations on chromosomal interactions is one of the critical aspects of genomics research. Traditional experimental approaches, including Hi-C technology, allow detailed investigation of genome interactions. However, they are usually costly, require specialized equipment, are time-consuming, and yield extensive data that requires further analysis. Furthermore, traditional approaches have difficulties analyzing the non-linear structure of the data. Thus, computational methods such as machine learning could be used to analyze genomic data. Machine learning offers an efficient and rapid solution to studying genomic data. Using information about interactions and the structure of the genome obtained via machine learning, connections between specific genome variations and structural features of the genome could be established.

II. RELATED WORK

The spatial organization of the genome and its impact on gene expression are key concepts in biology. A variety of experimental methods, including Hi-C, have been developed to examine communication between various parts of the genome. These studies demonstrate the importance of three-dimensional genome organization in controlling gene expression and maintaining cellular homeostasis.

At present, chromatin interactions are primarily analyzed using statistical and computational approaches. While these methods provide valuable insights, they do not account for all associations between the genome's architecture and gene expression regulation and are not designed to detect genetic factors that may influence genome-wide chromatin interactions.

With advancements in machine learning, the investigation of the genome's organization is now possible in a more comprehensive manner. Certain models can predict the chromatin interaction map and classify genomic sequences based on their structural properties. However, there are relatively few studies that explore the effects of genetic variations on genome-wide chromatin interactions. The present research proposal stands out from previous studies due to its unique approach.

III. SYSTEM ARCHITECTURE

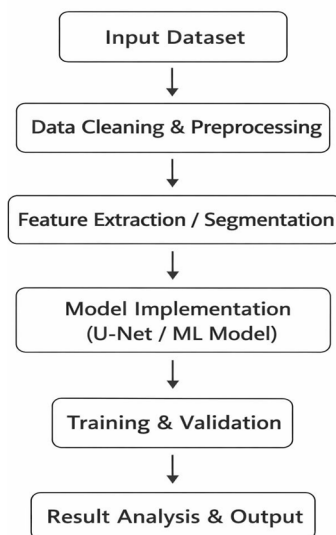
The spatial organization of the genome and its impact on gene expression are key concepts in biology. A variety of experimental methods, including Hi-C, have been developed to examine communication between various parts of the genome. These studies demonstrate the importance of three-dimensional genome organization in controlling gene expression and maintaining cellular homeostasis.

At present, chromatin interactions are primarily analyzed using statistical and computational approaches. While these methods provide valuable insights, they do not account for all associations between the genome's architecture and gene expression regulation and are not designed to detect genetic factors that may influence genome-wide chromatin interactions.

With advancements in machine learning, the investigation of the genome's organization is now possible in a more comprehensive manner. Certain models can predict the chromatin interaction map and classify genomic sequences based on their structural properties. However, there are relatively few studies that explore the effects of genetic variations on genome-wide chromatin interactions.

The present research proposal stands out from previous studies due to its unique approach.

Methodology



IV. EMERGENCY CONSCIOUS DESIGN & ACCESS CONTROL

The structure of the genome is analyzed based on chromatin interaction data to understand how genetic variants influence genome organization. The analysis focuses on interaction frequency, genomic distance between interacting regions, and the positional context of variants relative to regulatory elements. These factors provide insight into the structural characteristics of the genome and help in identifying patterns associated with structural disruption.

The proposed design emphasizes the importance of feature-based analysis, which allows the system to efficiently process high-dimensional genomic data. By focusing on relevant structural features, the system improves prediction accuracy and enables better understanding of genome organization.

A. Structure Features Analysis

The structure of the genome is analyzed depending on the characteristics of interactions obtained from chromatin interaction data. It relates to the interaction frequency, the distance between the interactions in the genomic sequence, and the localization of variants compared to the regulatory regions. The analysis of the mentioned features makes it possible to highlight the design features of structure that should be considered.

The reduction of the genomic sequence size in order to determine structural features makes it possible to increase the accuracy of machine learning models and recognize disruptive patterns in the genome structure.

B. The Variant Classification Algorithm

It is expected that the prediction of the effect of a particular variant on the structure of the genome based on data obtained from the training set leads to classification of the variant. In other words, classification of the variant becomes possible.

This approach can help predict and classify the variants quickly and accurately. Variants are classified as stable, borderline, and structurally disruptive.

C. Structure Disruption Detection

By using the suggested algorithm to detect disruptive variants, it becomes possible to identify regions in the genome that experience significant changes in chromatin interactions. This understanding helps in analyzing gene expression abnormalities and disease development processes.

The detection algorithm is essential for identifying critical genomic regions and improving the understanding of how structural variations contribute to diseases.

D. Design Scope

The design scope for the proposed model is confined to predicting the impact of the structure of the genome in an efficient manner through chromosomal interaction and machine learning approaches. To minimize the dependence on laboratory experimentation, the proposed prediction model will be capable of handling large amounts of data at affordable costs and less time.

Other features of the genome such as methylation and modification of histones can be incorporated into this model in order to make the prediction more efficient. The machine learning algorithms used in the prediction model can be substituted with deep learning algorithms.

V. IMPLEMENTATION DETAILS

The proposed system is implemented using Python due to its efficiency in handling and analyzing large-scale genomic data. Python provides a flexible and powerful environment for developing machine learning models and processing chromatin interaction datasets. Libraries such as NumPy and Pandas are used for data preprocessing, including cleaning, transformation, and organization of the dataset to ensure consistency and accuracy.

Machine learning algorithms are implemented using the Scikit-learn library, which provides robust tools for classification and evaluation. Models such as Support Vector Machine and Random Forest are used to analyze extracted genomic features and classify genetic variants based on their structural impact. These models are trained using relevant features and evaluated using performance metrics such as accuracy and precision to ensure reliable predictions. The system also includes a user interface that allows users to input genomic data and obtain prediction results. The interface is designed to be simple and user-friendly, enabling efficient interaction with the system. The output generated provides clear classification of variants as stable, borderline, or structurally disruptive, ensuring that the results are easy to interpret and useful for further analysis.

VI. RESULTS AND DISCUSSION

To estimate the suggested method's effectiveness, genomic interactions data related to the structure of chromatin were used to assess the algorithm's ability to detect potential disruptions. For example, it was found that the presented algorithm is capable of handling information related to such biologically important parameters as the distance between interacting chromatin elements, the frequency of their contacts, chromatin accessibility, density of genetic variants, and loop strength. Consequently, using the information above for analysis in conjunction with the machine learning model, one can understand the three-dimensional structure of the genome better.

Specifically, the presented genetic variants are divided into three categories depending on their influence on chromatin's structure. First, stable genetic variants do not have a significant impact on the structure of chromatin and, therefore, can hardly cause any changes in gene regulation mechanisms. The next category is represented by borderline genetic variants, which seem to affect the structure of the genome. This group deserves additional research because the variants' behavior may change with time, and disruptions may appear. Structurally disruptive genetic variants can negatively affect the structure of the genome and disrupt the gene regulation mechanism, leading to numerous types of genetic diseases.

Overall, one can conclude that the conducted analysis proves the efficiency of the machine learning model proposed in detecting the effect of various genetic variants on the structure of the genome. Additionally, due to the implemented categorization strategy, it is possible to determine which cases need further investigation. In conclusion, the created system represents an important achievement in the field of genomic research and may be applied in the future in exploring potential genetic diseases.

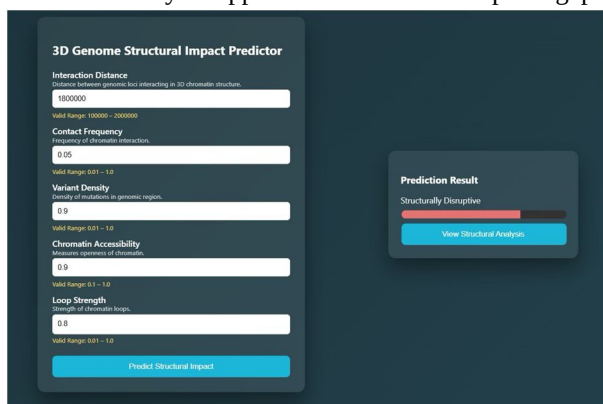


Fig. 1. The interface of the 3D Genome Structural Impact Predictor

Figure 1 shows the input interface of the 3D Genome Structural Impact Predictor, where users enter the necessary parameters in regards to the chromatin such as distance between interactions, the frequency of contact, the density of variation, chromatin accessibility and loop strength. To make accurate predictions, each parameter is limited to biologically sound values. After the inputs are typed in the system works with the data, applying a machine learning model, and produces a prediction of whether the genomic region is structurally disruptive. The outcome is presented in a graphic format with a status bar and the ability to have an in-depth view of the structural analysis, which makes the interface user-friendly and efficient.

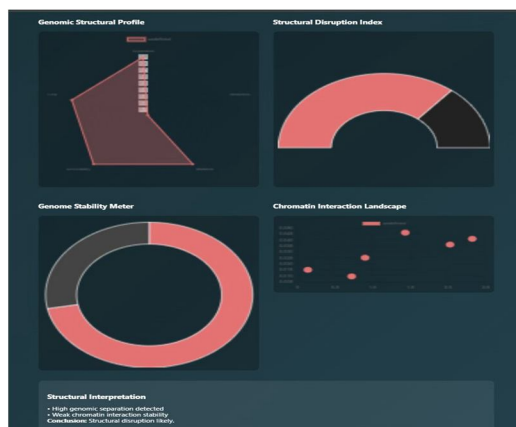


Fig. 2. Structural Analysis and Visualization Dashboard

Figure 2 shows the output visualization dashboard of the system which gives a detailed analysis of genomic structural impact. It contains several visual elements like a radar chart of genomic structural profile, a gauge of structural disruption index, a meter of genome stability and a plot of chromatin interaction landscape. All these visualizations are useful in comprehending the contribution of the various parameters to the prediction.

Also, a structural interpretation summary is provided with the highlights of the main findings, genomic separation, and interaction stability, and the conclusion on whether structural disruption is probable. The dashboard is more interpretable, as it converts the complicated model outputs into comprehensible and valuable understanding.

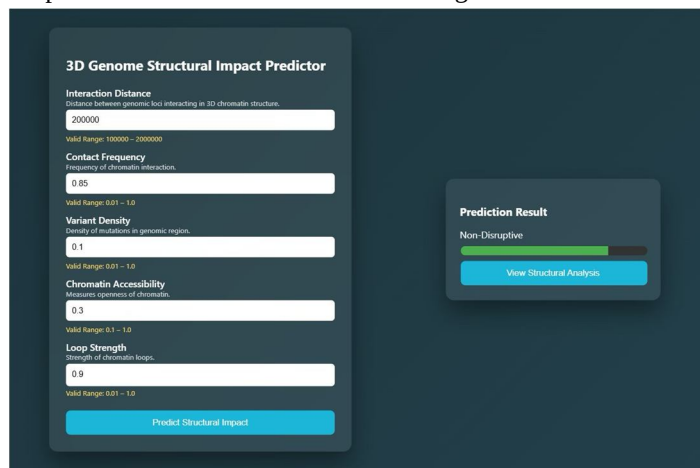


Fig. 3. Output Non-Disruptive Prediction Interface.

Figure 3 shows the input interface of the 3D Genome Structural Impact Predictor to which parameters parameters are chosen including interaction distance, contact frequency, variant density, chromatin accessibility, and loop strength to lead to a non-disruptive prediction. The inputs are fed into the system by the trained model and the result appears in the form of a non-disruptive, which is depicted with a visual representation of a green progress indicator. This means that structural integrity of the genomic region is upheld in the given conditions. The interface is clear and easy to use with the users being able to quickly understand whether the genomic configuration is stable or not.

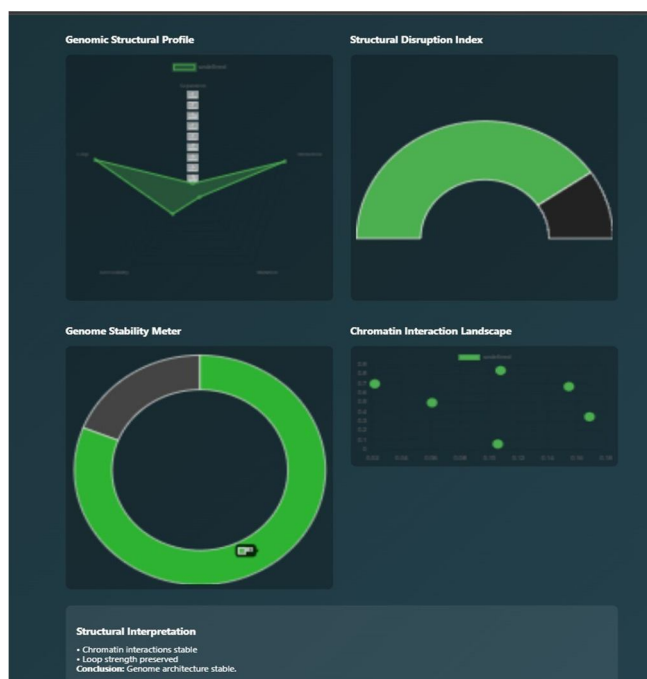
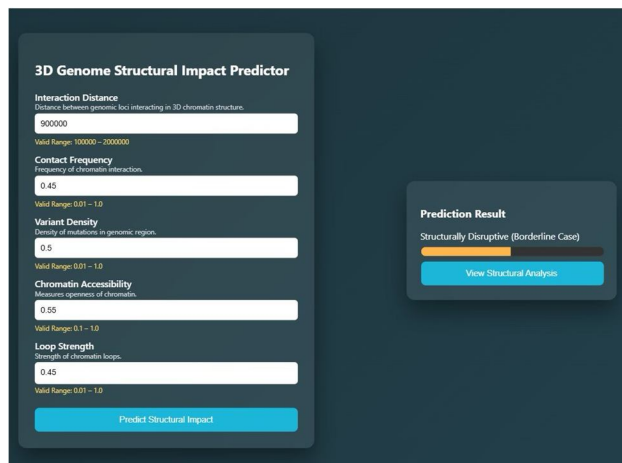


Fig. 4. Structural Stability Visualization Dashboard

The visualization dashboard for the case of non-disruptive prediction is presented in Figure 4 below. Genomic structures radar graph is balanced with equal contributions by parameters whereas the disruption meter for genomic structures shows low disruption levels. The genome stability meter is shown to have high levels of genome structures stability. The chromatin interaction meter shows that there are uniform interactions. Additionally, based on the genomic structure interpretation report, there are stable chromatin interactions along with stable loop strengths; therefore, genome structure is considered to be stable.



3D Genome Structural Impact Predictor

Interaction Distance
Distance between genomic loci interacting in 3D chromatin structure
900000
Valid Range: 100000 - 2000000

Contact Frequency
Frequency of chromatin interaction
0.45
Valid Range: 0.01 - 1.0

Variant Density
Density of mutations in genomic region
0.5
Valid Range: 0.01 - 1.0

Chromatin Accessibility
Measurements of chromatin accessibility
0.55
Valid Range: 0.1 - 1.0

Loop Strength
Strength of chromatin loops
0.45
Valid Range: 0.01 - 1.0

Prediction Result
Structurally Disruptive (Borderline Case)
View Structural Analysis

Predict Structural Impact

Fig. 5 Borderline Structural Disruption Prediction Interface

As shown in Figure 5, the prediction interface shows the case of a borderline structure for structural disruption because of the selected characteristics of the chromatin. The features, which include average interaction distance, contact frequency, number of variants, chromatin availability, and loop stability, are introduced to the machine learning model to get an output of “Structurally Disruptive (Borderline Case).” This output is illustrated by the middle (yellow/orange) bar, which means that the region of the genome is neither stable nor disruptive. This helps the user identify unstable regions that need further investigation.

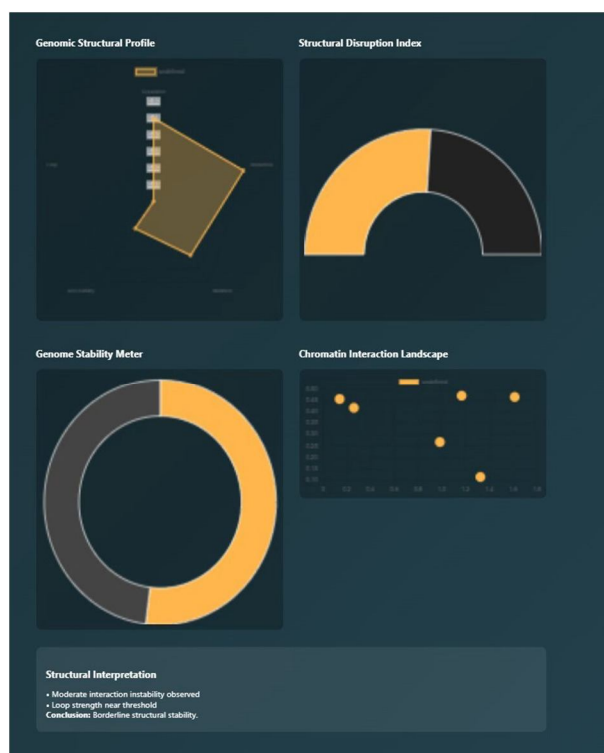


Fig. 6. Borderline Structural Analysis Visualization Dashboard

The visualization dashboard of the borderline prediction scenario has been shown in Figure 6, showing details of genomic structure. In the case of the genomic structure features profiling meter via radar plot, there is imbalance in the contribution of parameters; however, the structural disruption index meter shows moderate levels of disruption in genomic structural features. According to the genome stability meter, it shows balance between stable and unstable genomic structure, whereas in chromosome interactions landscape, variations in interactions have been observed. In addition, the summary of genomic structural interpretation suggests interaction instability and loop strength near threshold; thus, it reflects borderline state of genomic structure.

VII. CONSTRAINTS

The design and evaluation of the proposed system are carried out in a controlled laboratory environment using pre-existing chromatin interaction datasets. While this setup allows effective testing of the model, real-world challenges such as scalability and interoperability with diverse genomic data repositories have not been fully addressed. These factors may affect the system's performance when applied to large-scale or heterogeneous datasets. The proposed algorithm relies on a selected set of genomic features to analyze chromatin interactions and predict structural disruptions. However, it may not capture all aspects of chromatin organization, as genome structure is highly complex and influenced by multiple biological factors. The absence of multi-omics data integration, such as epigenomic and transcriptomic information, limits the comprehensiveness of the analysis. Furthermore, the system requires additional validation before it can be applied in clinical settings. Extensive testing with real-world datasets and validation under different biological conditions are necessary to ensure reliability, accuracy, and practical usability of the proposed approach.

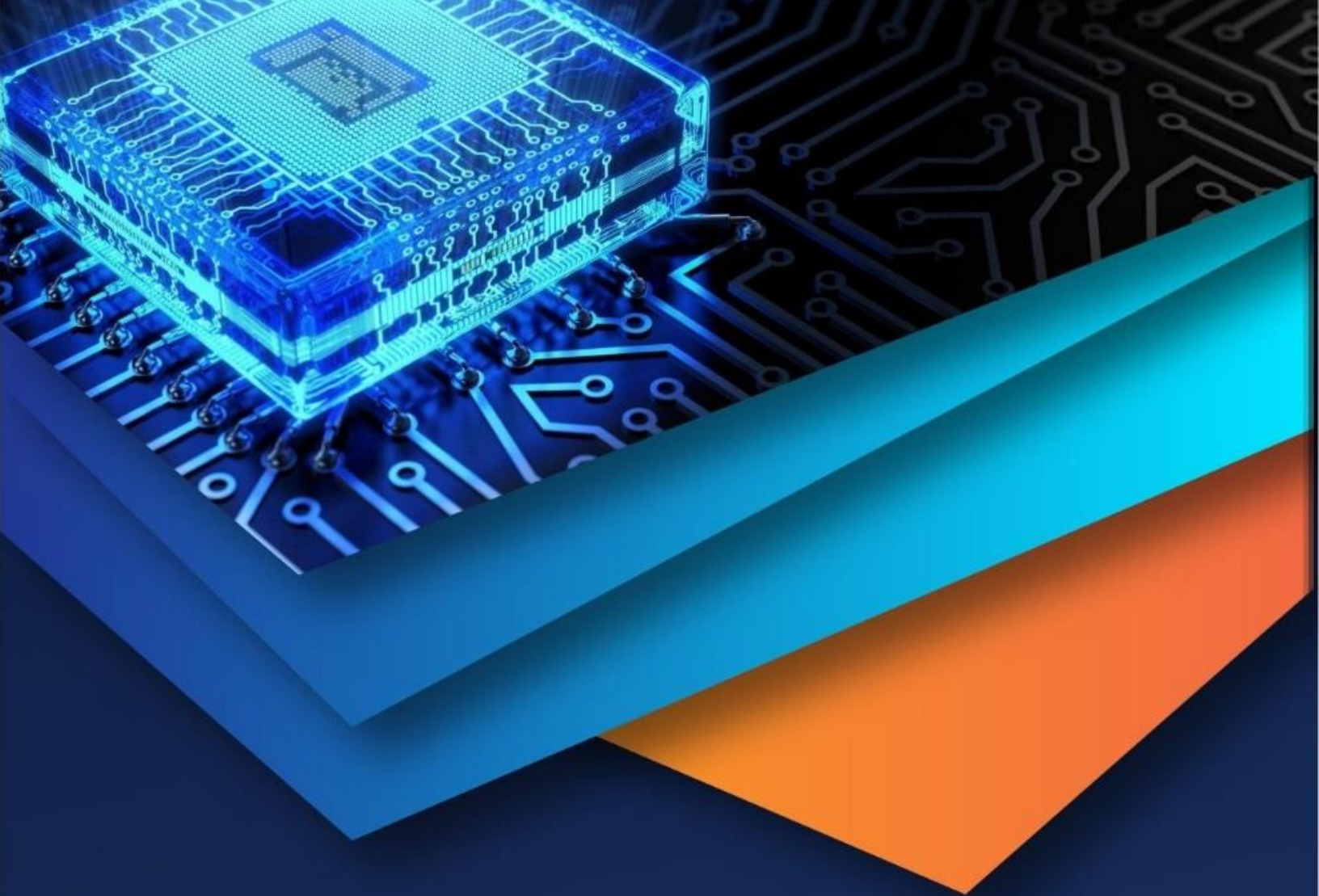
VIII. CONCLUSION AND FUTURE DIRECTIONS

The present work introduces a machine learning-based methodology for analyzing chromatin interaction networks and understanding structural disruptions caused by genetic variants. The proposed approach enables efficient analysis of genomic data without relying heavily on traditional laboratory methods, thereby reducing time and computational effort. By utilizing chromatin interaction features, the system provides a structured framework for studying genome organization.

The results demonstrate that machine learning algorithms can be effectively applied to detect structural disruptions and gain insights into genome architecture. This approach supports the identification of high-impact genetic variants and contributes to a better understanding of disease mechanisms. Consequently, it plays an important role in advancing precision medicine by enabling more accurate analysis of genomic data. Future research can focus on enhancing the proposed methodology by incorporating advanced deep learning techniques and integrating multiomics datasets. These improvements can increase prediction accuracy and provide a more comprehensive understanding of genome structure. Additionally, extending the system for real-time analysis and clinical applications can further improve its practical significance.

REFERENCES

- [1] A. Johnson, P. Liu, and T. Brown, Deep Learning Prediction of 3D Chromatin Structure from Genomic Signals, *Nature Biotechnology*, Vol. 41, 2023.
- [2] Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M. J. L., "Biopython: freely available Python tools for computational molecular biology and bioinformatics," *Bioinformatics*, vol. 25, no. 11, pp. 1422–1423, 2009.
- [3] Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., and Ren, B., "Topological domains in mammalian genomes identified by analysis of chromatin interactions," *Nature*, vol. 485, no. 7398, pp. 376–380, 2012.
- [4] H. Kim, S. Patel, and M. Park, Chroma Fold: Predicting 3D Chromatin Contact Maps Using Deep Learning, *Nature Communications*, Vol. 15, 2024.
- [5] Hunter, J. D., "Matplotlib: A 2D graphics environment," *Computing in Science Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
- [6] J. Schreiber, W. Stafford Noble, and K. Bilmes, Inferring 3D Genome Organization from Chromatin Interaction Data, *Bioinformatics*, Vol. 36, 2020.
- [7] L. Garcia, R. Thompson, and K. Singh, Machine Learning Reveals Diversity of Chromatin Interaction Patterns Across Cell Types, *Molecular Biology and Evolution*, Vol. 41, 2024.
- [8] Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., and Dekker, J., "Comprehensive mapping of long-range interactions reveals folding principles of the human genome," *Science*, vol. 326, no. 5950, pp. 289–293, 2009.
- [10] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [11] Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., and Aiden, E. L., "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping," *Cell*, vol. 159, no. 7, pp. 1665–1680, 2014.
- [12] X. Zhang, Y. Li, and J. Wang, Prediction of Cancer-Specific 3D Genome Organization Using Machine Learning, *Nature Communications Medicine*, Vol. 5, 2024.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)