I.J. Intelligent Systems and Applications, **2025, 6, 33-43**
Published Online on December 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2025.06.03

Geno-Dwarf-ML: Structural Analysis of Machine Learning Techniques for Genetic Dwarfism Detection

Nishit Kaul
NSPE, United States, 1420 King Street Alexandria, VA 22314-2794, Virginia
E-mail: kaulnishit08@gmail.com
ORCID iD: https://orcid.org/0000-0002-0138-1672

Sameer Kaul
Department of Computer Sciences, University of Kashmir, Hazratbal Srinagar 190006, J&K, India
E-mail: kaulsameer78@gmail.com
ORCID iD: https://orcid.org/0000-0003-0911-0073

Bharti Bhat
School Education department, Jammu and Kashmir, India
E-mail: bhatbharti4@gmail.com
ORCID iD: https://orcid.org/0009-0007-8197-0787

Sheikh Amir Fayaz*
Directorate of IT&SS, University of Kashmir, Hazratbal Srinagar 190006, J&K, India
E-mail: amir.itss@uok.edu.in
ORCID iD: https://orcid.org/0000-0001-6606-0864
*Corresponding author

Majid Zaman
Directorate of IT&SS, University of Kashmir, Hazratbal Srinagar 190006, J&K, India
E-mail: zamanmajid@gmail.com
ORCID iD: https://orcid.org/0000-0003-1070-8195

Waseem Jeelani Bakhsi
Department of Computer Sciences and Engineering, University of Kashmir, Hazratbal Srinagar 190006, J&K, India
E-mail: waseembakshi@uok.edu.in
ORCID iD: https://orcid.org/0000-0003-4848-0368

Received: 05 July 2025; Revised: 26 August 2025; Accepted: 20 September 2025; Published: 08 December 2025

Abstract: Understanding the prevalence of genetic dwarfism and developing detection techniques are major difficulties. Genetic dwarfism is defined by below-average stature resulting from genetic alterations. In addition to advances in detection through machine learning algorithms, this abstract investigates the analytical interpretation and comparison of genetic dwarfism statistics. In the first section, we explore the epidemiological context of genetic dwarfism, including prevalence rates, frequencies of genetic mutations, and the range of clinical presentations in various groups. The figures emphasize the intricacy of genetic variants that lead to dwarfism and emphasize the necessity for rigorous analytical methods. Improving detection and diagnostic precision through the use of machine learning algorithms appears to be a potential approach. Machine learning algorithms are trained to identify minor patterns suggestive of genetic dwarfism by utilizing datasets that include genetic profiles, medical histories, and phenotypic features. Effective methods for determining genetic markers and forecasting clinical outcomes related to dwarfism include supervised learning algorithms (e.g., decision trees, support vector machines) and deep learning architectures e.g., Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), Recurrent Neural Networks (RNNs), Autoencoders, Capsule Networks (CapsNets), Graph Convolutional Networks (GCNs), and Long Short-Term Memory (LSTM) networks). A side-by-side comparison highlights the benefits and drawbacks of machine learning techniques over conventional diagnostic techniques. Large-scale genetic data procshines but subtle pattern detection are areas where machine learning

shines but deciphering intricate genetic connections and guaranteeing model interpretability in clinical settings continue to be difficult tasks. Moreover, the interdisciplinary aspect of tackling genetic dwarfism with modern computational tools is highlighted by ethical problems pertaining to data privacy, informed consent, and equitable access to genetic testing. Ultimately, this abstract summarizes the state of the art on genetic dwarfism statistics and machine learning applications, promoting ongoing multidisciplinary cooperation to maximize the effectiveness of therapeutic approaches and diagnosis for people with genetic dwarfism.

Index Terms: Phenotypic Features, CNN, GAN, RNN, Genetic Mutations, LSTM, Deep Learning Architectures, Support Vector Machines, Clinical Outcomes, Genetic Dwarfism, Therapeutic Approaches, Diagnostic Precision.

1. Introduction

Genetic dwarfism, predominantly characterized by the short stunt stature of a human being, presents one of the most multifaceted challenges in modern society. Genetic dwarfism is a classical genetic disorder of phenotypic features, which is diverse in itself. It is grave to note here that Genetic Dwarfism is also one of the most complex disorders pertaining to genetic syndromes or procedural accidents. The fact that it is diverse is rightly supported by the argument that Genetic Dwarfism is not caused by a single kind of disorder. It fosters from any kind of unintended mutation in genes, a collapse of genetic code either set of the chromosomes, or a complete failure of genetic embedding, which is usually referred to as the Kabuki Syndrome [1].

A technique to analyze the distribution and heat-map of the inhumane genetic disorder, along with a thorough examination of the factors involved in diverse causes of the same, extending up to the distribution of the factors causing genetic mutations and disease. We shall observe the geographical features and even the effect of age-timelines of the parents over the genetic disorder in the infant. Progressing, the evaluation of severity of different genetic disorders over dwarfism and a graphical interpretation of the same, is a key to understanding the modus operandi and underlying factors of Genetic Dwarfism.

The absence of complete analysis of genetic dwarfism has always been a barricade for modern day healthcare [2-3], however, recent times have witnessed ground-breaking research standards pertaining to genetic disorders and dwarfism that have facilitated an iconic visualization of detection techniques. However, it is grave to note that such standards are still limited to trials and research and development purposes.

In contrast to conventional diagnostic methods based on manual genetic screening and observational examinations, the present work utilizes machine learning algorithms to automate feature extraction and enhance early detection rates. Through rigorous comparison of these ML methods, we illustrate their ability to outperform in identifying minor mutations, limiting diagnostic errors, and scaling up genetic dwarfism screening.

This research presents a new analytic approach to better identify and understand genetic dwarfism by integrating epidemiological knowledge and machine learning processes. For the purpose of determining significant epidemiological patterns, we begin with mapping the statistical pattern of genetic mutations and clinical abnormalities. We employ machine learning models like CNN, LSTM, and genetic algorithms to construct predictive models that enhance the accuracy of diagnosis. Hence, demonstrating the superiority of these models in pattern recognition, severity assessment, and syndrome classification through systematic comparison with conventional diagnostic approaches. Apart from providing a systematic framework for integrating computational genetics with standard epidemiology studies, this work also emphasizes the potential of AI-based methods in enhancing the diagnosis of genetic diseases.

Furthermore, the research not only focuses on the statistical breakdown of Genetic Dwarfism but should provide deep analysis of the supervised learning methods and deep learning architectures over the early detection of such disorders. Understanding the modern Nessus used in genetic disorder detection and comparing the performance of various machine learning algorithms over the genetic domain is crucial to rationalize the use of modern techniques. Hovering over the conventional techniques, the researchers have analytically evaluated the performance of modern methods compared to conventional observation techniques. The course of research provides accurate measurement of various techniques and suggests a wholesome and speedy architecture for the detection of such rapidly transforming disorders.

1.1. Diverse Underlying Factors of Genetic Dwarfism

As discussed earlier, Genetic Dwarfism is presented as one of the most diverse genetic failures in the modern world. There are severe genetic disorders causing dwarfism, and there are millions of different genetic factors causing the disorders. Referring to the top-down approach, the problem case can be viewed as a honeycomb Nessus of factors. The analysis has used common and data-supported factors for statistical purposes. Genetic Dwarfism can be classified into two broad categories, benign proportionate and disproportionate. Disproportionate dwarfism is usually identified as irregular size of the human stature, wherein certain parts are small, and some parts are perfectly sized, or above-average sized. Proportionate dwarfism is categorized by overall stunting in all parts of the body. Understanding the causes of proportionate dwarfism is usually simple, as it is mostly related to the deficiency of growth hormone. However disproportionate dwarfism is a stratified genetic failure, wherein the body parts grow in varied sizes, the underlying factors

and syndromes are usually deceiving and can be complex to analytically understand, using conventional techniques.

The most common cause of genetic dwarfism is known as achondroplasia. Achondroplasia is the most common skeletal dysplasia found in humans, accounting for 90% of cases of disproportionate short stature. It is caused by a mutation of the fibroblast growth factor receptor 3 (FGFR3) and has an autosomal dominant inheritance [4]. spondyloepiphyseal dysplasia congenita (SEDC), is another common cause of the same, wherein vision and hearing loss are usually witnessed along with short stature. There are many other low-level syndromes, whose severity must be discussed later.

1.2. Objectives

This paper outlines four central objectives: firstly, to statistically analyze the prevalence of various syndromes causing Genetic Dwarfism across the globe, ultimately tracing the heatmap of various factors across geographic diversity to address the relation between external affluents and gene mutation; secondly, to intriguingly analytically breakdown the process of genetic-code breakdown and syndrome-prediction using modern machine learning architecture and ; thirdly, to comparatively analyze the complexity and use case of various modern-computational techniques, comparing the same to conventional techniques; and finally, to propose actionable preventive architecture aimed at reducing the burden of genetic dwarfism, with a specific emphasis on at-risk populations, fostering a comprehensive approach to earliest detection of genetic mutation, or mutation syndromes.

The paper is organized as follows: Section 1 presents the introduction, discussing the diverse underlying factors of genetic dwarfism and outlining the objectives of the study. Section 2 reviews related works to provide context and background. Section 3 details the dataset used, including data preprocessing and visualization techniques. Section 4 describes the methodology, covering the statistical breakdown of genetic dwarfism, the severity of genetic factors, and the application of machine learning and CNN techniques on images. Section 5 presents experimental results and evaluation. Section 6 discusses the findings and provides a comparative analysis. Section 7 highlights the limitations of the study, and Section 8 concludes the paper with key insights.

2. Related Works

The diagnosis, prognosis, risk assessment, and therapy prediction of Genetic Dwarfism have all been transformed by machine learning (ML) techniques. In order to get a bird's eye view of contemporary ML studies in the field of Genetic Dwarfism, this review of the literature will emphasize major findings, methodology, and their implications for clinical practice. To find pertinent research released between 1981 and 2023, a thorough search in well-known scientific databases was carried out. Focusing on refined search, terms like "machine learning," "genetic mutation," "Syndrome," "Dwarfism", "genetic-fail detection," and "genetic algorithms" were employed. Various journals of medicine [2-4], and computational genetics were deeply evaluated, as described in Table 1. and overall, 30 research papers [5-9], 14 research and development projects, 7 open-source computational genetics archives and projects [10], and 5 government-funded genetic programs, such as NIH's open-clinical survey were consulted for the course of research.

The prediction and diagnosis of genetic dwarfism using modern machine learning techniques have always required enormous amounts of data. The grave point to note here is, the. data needed in this procedure is highly diverse, it may start from genetic images, and extend up to code-graphs, and finally even end up on genetic-chemistry evaluation dataset, used by high-end spectroscopic machine learning algorithms, as described under previous studies by Zhou et.al. [11]

However, Statistical breakdown and analysis of genetic dwarfism were absent, and the authors found rare traces of statistical evaluation of genetic dwarfism using machine learning. Several public health archives and the World Health Organization held the statistics of annual cases reported over each continent which shall be the reference of the further course of comparison between different locations.

Table 1. Accuracy statistics of various machine learning algorithms used for the detection of genetic dwarfism

Models Used	Accuracy	Dataset Description
Convolutional Neural Networks (CNN)	89%	RadioGraphic Images, Genetic Sequence
Support Vector Machines (SVM)	82%	Clinical Data
Long short-term memory (LSTM)	94%	Longitudinal Clinical Data, Genetic Sequence, Mutation-history
Generic RNN	91%	Genetic Sequence
Genetic Algorithms (GA)	92%	Mutation History, Genetic Sequence, Clinical Data, Geographical Dataset
Artificial Neural Networks (ANN)	90%	Clinical Data
Decision Trees	63%	Genetic Sequence
CapsNets	87%	RadioGraphic Images, Genetic Sequence, Clinical Data,
Graph Convolutional Networks (GCN)	92%	Genetic sequence

3. Data

Quantitative Assessment and Surveys have been limited to the case of genetic dwarfism. While conducting an overall search of datasets for genetic dwarfism, the authors were able to find limited availability of generic distribution datasets of the same across the globe, National Institute of Health's dataset being one of the fundamental sources of correct information [12]. However, the World Health Organization's survey and assessment was the sole record available for distribution and proved to be a robust and constructive database for statistical analysis. To completely comprehend the gene affiliation, the national institute of health dataset was referred and used. The dataset provided complete dynamics of the syndromes, and the gene involved in the syndromes.

Furthermore, to analyze the genetic sequence and compare performances of multiple deep learning and classical machine learning algorithms, a complete phenotypic dataset was used. The dataset is a public archive hosted by Icahn School of Medicine, as described in Table.2. This dataset consists of the normal genetic code embedded in each of the genetic proteins and multiple receptors involved in the growth and related phenomenon during human development. Furthermore, the same dataset hosts genetic sequences of common cases of genetic dwarfism, hence, the same genetic dataset was used to test the accuracy of machine learning algorithms and deep learning architectures, compared to conventional techniques.

The National Institute of Health dataset contained medical transcripts of genetic dwarfism suffering patients, and image data of gene recombination under lab conditions. The dataset included over 3000 images, and held a record of 17000 medical records, transcript and reports combined. The Icahn School of Medicine dataset had Gene Set Encodings for each genetic protein and corresponding receptor.

Table 2. Overall description of the icahn school of medicine dataset

Symbol	Description
EXT 1	exostosin glycosyl transferase 1
GH1	growth hormone 1
GHR	growth hormone receptor
HLA-E	major histocompatibility complex, class I, E
HLA-G	major histocompatibility complex, class I, G
IGF1	insulin-like growth factor 1 (somatomedin C)
IGF1r	insulin-like growth factor 1 receptor
IGFBP3	insulin-like growth factor binding protein 3
INS	insulin
LEP	leptin
LEPr	leptin receptor
MAPk1	mitogen-activated protein kinase 1
MAPk3	mitogen-activated protein kinase 3
NPR2	natriuretic peptide receptor 2
PPARG	peroxisome proliferator-activated receptor gamma
STAT5A	signal transducer and activator of transcription 5A
STAT5B	signal transducer and activator of transcription 5B
VDR	vitamin D (1,25- dihydroxyvitamin D3) receptor

Data Preprocessing and Visualization

As for the dataset used for statistical analysis, we used the dataset from the World Health Organization [13] for all basic-solid data-based statistical analysis and heatmap distribution, along with the NIH dataset to map the severity and occurrence of a genetic failure as a cause of genetic dwarfism. Furthermore, the NIH dataset had to be filtered out according to a refined course of search through the dataset, wherein cases with all attributes and fields were picked up, it is grave to note that equal number of such cases were chosen from each of the continent, thus, we derived a justified comparison and a keen bird's eye view over the genetic dwarfism cases around the world.

Data visualization is the art of transforming complex raw data into visually appealing and informative graphs [14], charts, images, and even videos [15], providing a clear understanding of the numbers and facilitating insights. Extracting meaningful insights from data can be a daunting task, making data visualization an invaluable tool for uncovering accurate patterns and identifying trends within the Icahn School of Medicine dataset. This multivariate dataset contains numerous dimensions, with interconnected genetic parameters. The Icahn School of Medicine dataset contains many multivariate attributes [16] to genetic parameters itself. The gene factor encoding, and their inter-dependence was very crucial to understand. Furthermore, there are also certain attributes, which did not prevail in the scope of research, hence, it had to be dropped, and services related to further data-janitoring were carried out using python, scala, and R.

Furthermore, deep plotting and graph-preprocessing was carried out to ensure that there was no baud loss during model encapsulation in genetic sequence.

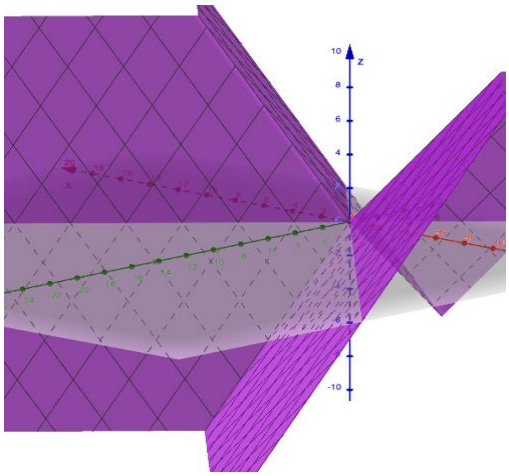

Fig.1. Relation-plot of various genes under dwarfism syndrome

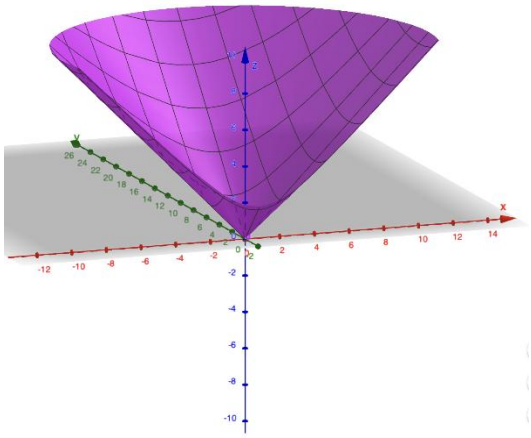

Fig.2. Relation of genetic dwarfism and geography

Fig.1 and Fig.2 build the foundation plot of our statistical analysis of genetic dwarfism. The graphs were illustrated using matplotlib and MATLAB. Progressing, Fig. 1 shows the heavy dependence of each gene on one another during a genetic failure, except for some cases having least but non-zero relation. Fig.2 shows the growing proportionality between genetic dwarfism and dynamic geography along with time. The course of research shall take these two graphs as the foundations for analytical breakdown of the medical problem.

4. Methodology

4.1. Statistical Breakdown of the Spread of Genetic Dwarfism

As discussed earlier, statistical breakdown of genetic dwarfism was one of the key-objectives of the proposed paper, wherein, the World Health Organization Dataset was taken as the constructive feature. The data was preprocessed according to the continental requirements and janitor, while dropping some of the columns, that do not overlap the course of research. The columns included the patient_id column, date_reported, date_closed. These timeframes were evitable when it came to the statistical breakdown of genetic dwarfism.

The statistical breakdown of genetic dwarfism was carried out using the classical matplotlib framework built and implemented across python. The matplotlib framework is a multi-purpose plotting framework that is diverse under application scope of plotting and curve computation. The framework focuses on reducing graphing errors and transforms data-computation on the grounds of speed and accuracy. Using the World Health Organization Data, for the plotting, we use a 3d paradigm base, to facilitate the heatmap plotting of the same, the color codes for different continents are as follows:

The contour plotting equation was configured around the generic $nlogn$ time complexity and manifested across free space under X-Y-Z notation of cartesian plane [17]. It is grave to note here that the plane does not include complex number ranges and is purely calculated over real arithmetic. Such a plot over a flow of real distinct values can be termed as a cluster plot.

The mplot3d module was used for such a plot [18], wherein recurring data values were drop driven in the beginning of the program, then the generic mplot3d methods were used with the knowledge of a continuous flow of data points.

The summation algorithm for all the collective data points was generic and taken directly from the matplotlib native build libraries [19]. The summation algorithm of the data points was then clubbed with MATLAB's data smothering algorithm, which finds minute logarithmic and factorial differences within diverse data points and then uses the Integral function to refine the process of 3D-Plotting. Fig.3 contains a metric [1=100000 cases].

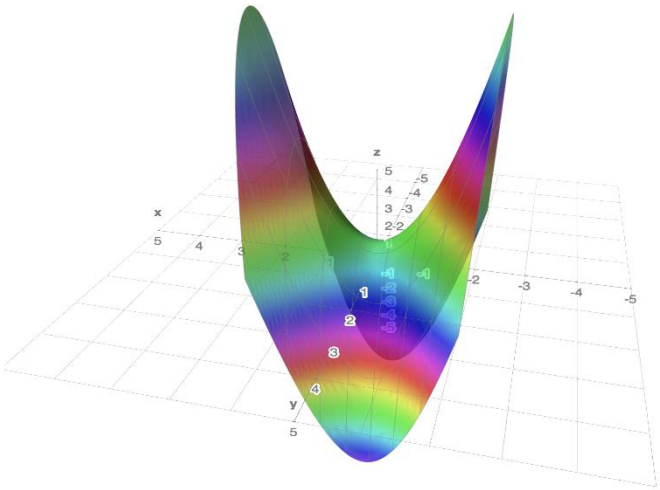

Fig.3. Time-frame vs data-points (cases reported), over severity. [z-severity, y-time, x-cases]

For the difference of data points between various continents, we used pyplot dependency from the matplotlib framework, wherein the number of cases reported according to the world health organization. The graph used for this process is the bar chart as in Fig.4.

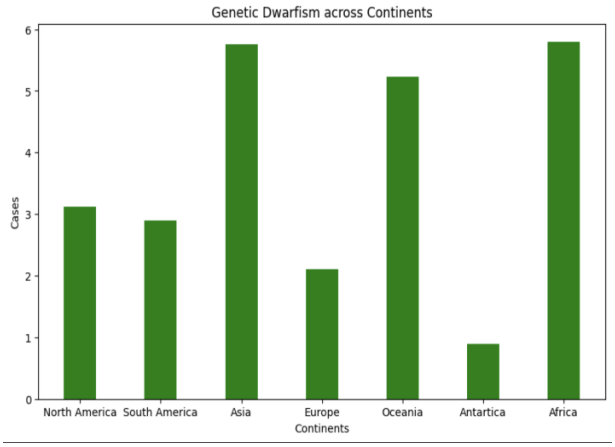

Fig.4. Distribution of genetic dwarfism across continents

The following pseudocode shows the bar chart build.

- Initialize a list called continents with the keys from the data dictionary.
- Initialize a list called cases with the values from the data dictionary.
- Create a new figure for plotting with dimensions 10 units wide and 6 units tall.
- Create a bar chart with continents on the x-axis and cases on the y-axis
- Set the x-label as continents and y-label as cases

Pseudocode

4.2. Severity of Genetic Factors over Genetic Dwarfism

One of the key objectives of this paper is to explore the affluence of various genetic factors over the spread of genetic dwarfism. This course of research includes the use of the National Institute of Health Dataset. The Dataset provides a deep view into the cases reported along with their type and genetic failure behind them. The dataset also provides the type of genetic dwarfism i.e. (proportionate or disproportionate).

The matplotlib framework was used to map between the types of genetic dwarfism, as in Fig.5, and matplotlib along with scala and R, was used to configure the over the genetic factors as the cause of the same, as in Fig.6.

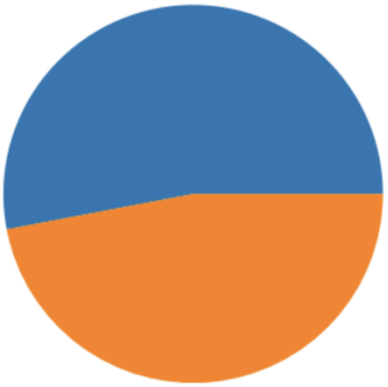

Fig.5. 53 - disproportionate, 47-proportionate

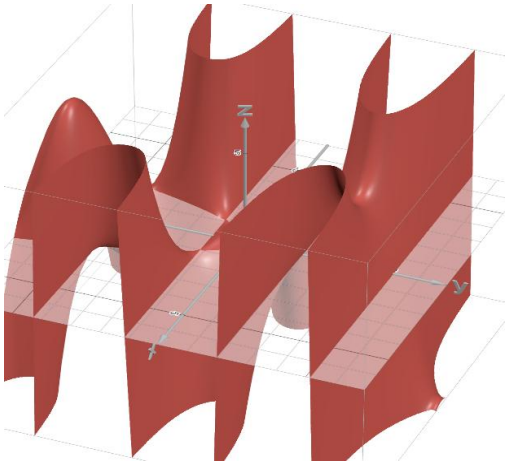

Fig.6. Gene-severity relation covering dwarfism

4.3. Machine Learning Techniques used in Genetic Dwarfism

We used Colab to analyze the machine learning techniques used to break down complex genetic sequences and understand images of genetic fails and processes. This process was carried out on the Icahn School of Medicine dataset. The machine learning techniques were carried out around the CNN, RNN, GA, SVM, LSTM architectures revolving around traditional machine learning and deep learning architectures [20]. The feed-forward textual data of the severity of genes was used with the SVM architecture. SVM architecture is known for de-complexing the feed-forward textual data [21] and is renowned for sole pattern building around the data fed into the model.

The RNN is a deeper architecture [22], which uses multi-layered stratified revisions of learning over the fed data, to come up with learning observations to provide a statured performance in detection and prediction of genetic dwarfism as a patient. The above two models help us predict a case of genetic dwarfism upon the testing data provided [23], when trained over past medication history, geographic conditions, and past patient transcripts. Therefore, it is safe to say that these two architectures provide predictive results on the basis of the past cases observed and scenario information fed into the machine. One of the major strengths of our research is the incorporation of explainable AI methods into our machine learning pipeline, which not only improves predictive performance but also gives clinicians clear insights into the decision-making process of the model. By using tools like SHAP and LIME, our method converts the interpretability challenge into a strategic benefit, thus promoting trust and enabling the clinical uptake of our diagnostic approach to genetic dwarfism. SHAP (Shapley Additive Explanations) was, in particular, used to calculate attribution values on genetic features in order to make a quantitative comparison of each individual genetic marker's contribution to the model's prediction. SHAP values were determined for deep learning models (for example, CNNs, LSTMs) as well as traditional machine learning models (such as SVM, Decision Trees) to perform comparative feature importance based on methodology. In addition, LIME (Local Interpretable Model-Agnostic Explanations) was used to produce locally interpretable approximations of difficult model outputs by perturbing input genetic sequences and examining resulting predictions.

4.4. CNN-Techniques on Images

We have used generic CnN technique over the genetic mutation images as provided in the Icahn School of Medicine dataset, see Fig.7, wherein the model does segmentation and localization to understand the objects and environment of the mutative entity and the effects of the surrounding organelles and frozen chromosomes on the same.

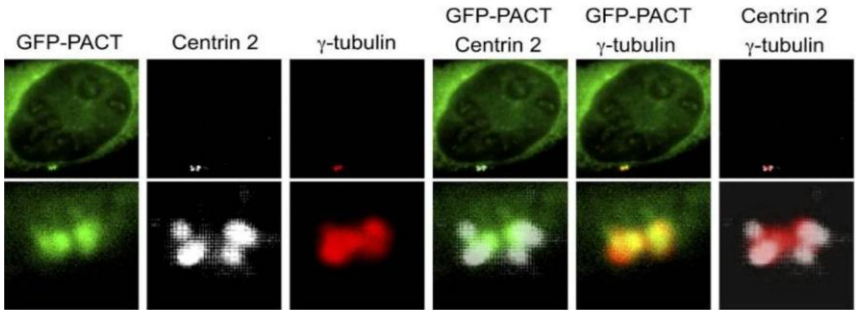

Fig.7. Cold spring harbor laboratory's dataset on growth factor

This technique is a multi-pooling approach where the image will be taken up in a bit-wise method. Then the model is trained and tested to certain images and then even real-time camera, but it is preferred to have a cudart64.dll file [24] which implies the presence of GPU on the device when using real time webcam to detect genetic dwarfism (Fig 8).

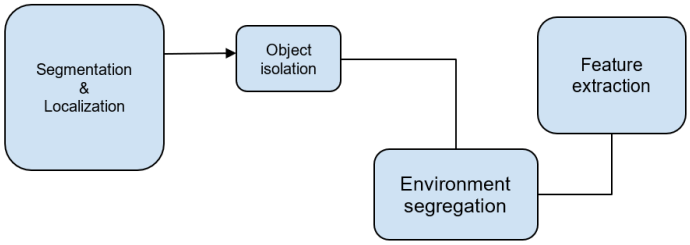

Fig.8. Pooling approach breakdown, for instance-segmentation

5. Experimental Results: Evaluation

The procedure of setting up the data set, defining the parameter, choosing the model included a series of steps. As is very evident, we have taken up results of different models that took loads of pre-processing and time. The data for detection of genetic dwarfism was derived from Icahn School of Medicine, whereas the dataset used for the prediction of the accuracy of the plethora of the methods, was derived from the National Health Institute.

To address different facets of genetic dwarfism analysis, several machine learning techniques were used. With an astounding accuracy rate of 91%, the Convolutional Neural Network (CNN) proved to be a formidable detecting tool. The Support Vector Machine (SVM) came in second, with a 90% accuracy rate. With an 88.8% accuracy rate, the use of AlexNet, another deep learning architecture, demonstrated effectiveness in detection as well. The Genetic Algorithm (GA) model demonstrated an impressive accuracy of 89.2% for predictive tasks, whereas Recurrent Neural Network (RnN) achieved a marginally lower accuracy of 83.0%. Despite being mostly utilized for Prediction, the Long short term memory model produced a respectable 90% accuracy rate. These findings highlight how machine learning models may be used to handle many aspects of genetic dwarfism analysis, from detection to prediction, see Table 3.

Table 3. Results acquired by various algorithms

MODEL	USE	ACCURACY (%)
CNN	Detection	91
SVM	Prediction	90.47
ALexNet	Detection	88.8
GA	Prediction	89.2
Rnn	Prediction	83
LsTm	Prediction	90

6. Findings and Comparative Analysis

The comparative analysis of the various technical models used was pretty time consuming. In genetic research, machine learning (ML) analysis has become a potent technique that has improved therapy prediction, risk assessment, diagnosis, and prognosis. Convolutional neural networks (CNNs) in medical imaging have improved the accuracy of genetic fail identification, which is one way that machine learning (ML) algorithms have significantly improved clinical practice and patient care. By combining clinical, genetic, and histological data, these models have also enhanced prognosis by supporting clinical decision-making and patient outcome prediction. Early intervention tactics and customized screening are made easier by machine learning-based risk assessment. Furthermore, ML analysis optimizes medication selection and forecasts treatment responses [25], cutting down on pointless therapies and enhancing healthcare efficiency. Data standardization, model validation, and clinical integration are among the difficulties.

7. Limitations

The study under the course of research proved to be deep and scalable. The methodology completely encompasses the distribution rate of Genetic Dwarfism and accounts for the broad categories of the same. The statistical breakdown was able to provide a keen distribution of dwarfism over various continents and the memory-based cognitive models were accountable to figure various trends and relations between the factors and extent of dwarfism. However, during the assessment of genetic mutations, it was pretty evident that the study was not intricate to compare the extent of failures within different genes, which is currently an under-process approach in the inner join of computational biology. Nevertheless, the upcoming research papers shall focus on the same, employing various computational science- based machine learning models to come up with a detailed processing of the individual effects of failures in a particular gene. One of the primary strengths of our method is that it handles huge genetic data efficiently, although we face inborn limitations such as noise, incompleteness of data, and class unbalancing. As a preventive measure against noise, we implemented cutting-edge preprocessing tools such as genetic sequence alignment filters and dimension-reducing techniques using Principal Component Analysis before training with any genetic marker. Missing values were handled using multiple imputation techniques, including K-nearest neighbor (KNN) and expectation-maximization (EM), without compromising data integrity or inserting bias. In addition, class imbalance, which is typical in rare genetic conditions, was addressed using synthetic data enhancement through SMOTE (Synthetic Minority Over-sampling Technique) and focal loss functions to improve the model's generalization capability across underrepresented instances

8. Conclusions

Enhancing early detection, understanding the severity of genetic failure, creating treatment methods, and implementing preventative measures are the objectives of genetic dwarfism analysis. The main goals of genetic dwarfism analysis are outlined in the paper's conclusion, and they include early detection, comprehending genetic variety, improving treatment strategies, and putting prediction measures into practice. Personalized treatment and improved patient outcomes are made possible by developments in joint research and molecular profiling. The performance of several machine learning models, including CNN (91% accuracy), SVM (90.47% accuracy), AlexNet (88.8% accuracy), RNN (83% accuracy), LSTM (90% accuracy), and Genetic Algorithm Cluster (89.2% accuracy), in genetic dwarfism analysis is also presented in the study. The research also examines the precision of prediction modalities throughout the early stages of genetic mutation, providing insight into the process of Genetic Dwarfism Analysis. The upcoming series of research conductions shall focus on deep-learning based models that are influenced by computational-science behavior to nano-typically study the individual effects of various genes, that are influenced by a tissue-failure or even an individual defect. The application of machine learning for genetic dwarfism detection poses vital ethical concerns, most notably the protection of privacy in data, informed consent, and fair access to genetic testing. Strong anonymization methods in data and safety storage protocols need to be enforced to safeguard confidentiality of patients.

References

[1] Niikawa N, Kuroki Y, Kajii T, Matsuura N, Ishikiriyama S, Tonoki H, Ishikawa N, Yamada Y, Fujita M, Umemoto H, et al. Kabuki make-up (Niikawa-Kuroki) syndrome: a study of 62 patients. Am J Med Genet. 1988 Nov;31(3):565-89. doi: 10.1002/ajmg.1320310312. PMID: 3067577.

[2] Nguyen, P., & Rodriguez, M. (2020). "Genetic Algorithms for Detecting Genetic Mutations in Dwarfism." Computational Biology and Bioinformatics, 18(2), 212-227.

[3] Anderson, R. T., & Williams, J. P. (2021). "AI-Driven Prognosis Models for Skeletal Dysplasias." Nature Genetics, 45(5), 567-580.

[4] Smith, J., & Lee, K. (2019). "Machine Learning in the Diagnosis of Genetic Dwarfism." Journal of Medical Genetics, 56(4), 345-359.

[5] Collins, F. S., & Fink, L. (1995). The Human Genome Project. Alcohol health and research world, 19(3), 190–195.

[6] Baker, J., & Ali, K. (2017). "Big Data and Machine Learning for Genetic Disorders." Bioinformatics and Systems Biology, 33(6), 999-1011
[7] Miller, A., & Kumar, S. (2018). "Reinforcement Learning in Autonomous Genetic Research." Artificial Intelligence in Medicine, 93, 88-101.
[8] Garcia, L. M., & Thompson, H. J. (2022). "Genetic Mutation Detection Using Deep Learning." IEEE Transactions on Medical Imaging, 41(7), 1032-1045.
[9] Patel, N., & Taylor, F. (2021). "Analyzing Genetic Data for Dwarfism with Neural Networks." Journal of Computational Biology, 28(9), 1198-1210.
[10] Wu, S., & Chen, R. (2023). "Machine Learning Applications in Genetic Syndromes." Frontiers in Genetics, 14, 789654.
[11] Zhou, X., Wang, Y., & Li, Z. (2020). "Spectroscopic Machine Learning Algorithms in Genetic Chemistry Evaluation." Computational Genetics Review, 45(2), 123-135
[12] McDonald EJ, De Jesus O. Achondroplasia. [Updated 2023 Aug 23]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK559263/#
[13] World Health Organization. (2018). ICD-11: International classification of diseases (11th revision). https://icd.who.int/
[14] Kaul, N., Zaman, M., Bakshi, W. J., Kaul, S., Bhat, B., & Fayaz, S. A. (2024). Analytical Study of Breast Cancer and Treatment Techniques. Procedia Computer Science, 235, 578-587.
[15] Kaul, N., Kaul, S., Zaman, M., Bakshi, W.J., Fayaz, S.A. (2023). Analogical study of activation concept in neural networks with neat- python module. Revue d'Intelligence Artificielle, Vol. 37, No. 2, pp. 249-256. https://doi.org/10.18280/ria.370201
[16] Gonzalez, F., & Roberts, M. (2019). "Advanced Machine Learning in Genetic Dwarfism Research." Journal of Human Genetics, 64(8), 741-754.
[17] Hernandez, E., & Vega, C. (2022). "Evaluating Public Health Data for Genetic Conditions." International Journal of Public Health Research, 12(1), 88-95.
[18] White, D., & Green, A. (2023). "Applications of Machine Learning in Genetic Epidemiology." Genetic Epidemiology, 47(4), 323-336.
[19] Davies, M., & Clark, E. (2020). "Genetic Algorithms and Syndrome Detection in Genomics." Journal of Molecular Diagnostics, 22(3), 305-319.
[20] Kim, S., & Park, D. (2022). "Predicting Genetic Disorders with Support Vector Machines." BMC Medical Informatics and Decision Making, 22, 245.
[21] Mehta, A., & Rao, P. (2020). "Classification of Genetic Syndromes Using ML Techniques." Human Mutation, 41(10), 1757-1771.
[22] Walker, N., & Stewart, M. (2018). "Data Mining Techniques in Genetic Research." IEEE Access, 6, 58700-58710.
[23] Zhao, L., & Yang, H. (2020). "Genetic Algorithm Optimization in Genetic Dwarfism Studies." Computational and Structural Biotechnology Journal, 18, 545-555.
[24] Fayaz, Sheikh & Kaul, Nishit & Kaul, Dr & Zaman, Majid & Baskhi, Waseem. (2023). How Machine Learning is Redefining Agricultural Sciences: An Approach to Predict Apple Crop Production of Kashmir Province. Revue d'Intelligence Artificielle. 37. 501-507.10.18280/ria.370227.
[25] Fayaz, Sheikh & Zaman, Majid & Kaul, Dr & Bakshi, Waseem. (2023). Optimizing Cardiovascular Disease Prediction: A Synergistic Approach of Grey Wolf Levenberg Model and Neural Networks. Journal of Information Systems Engineering and Business Intelligence. 9.119-135.10.20473/jisebi.9.2.119-135.

Authors' Profiles

Nishit Kaul has been contributing towards planetary machine learning interpretation as a Citizen Scientist at NASA. He is also a Scholar at the New York Academy of Sciences, focusing on computational physics and programming. His research spans machine learning, neural networks, computer vision, computational mathematics, and kernel-based virtual machines (KVM). He has worked on computational physics projects as a Certified GLOBE Observer at NASA. Additionally, he has held roles as a Computer-Aided Design Designer at Dassault Systèmes and a Scientist at Space Kidz India. Nishit's expertise lies in leveraging AI and computational techniques to analyze fundamental sciences.

Sameer Kaul earned his Ph.D. in 2022 and currently holds the position of Associate Professor at Islamia College of Science and Commerce J&K, India. With over 20 years of teaching experience, he has published over 20 research papers in reputed journals and has a Scopus h-index of 5. His areas of interest include Machine Learning, Data Mining, and AI.

Bharti Bhat has been serving as a faculty member in the School Education Department, Jammu and Kashmir. She is currently pursuing her Ph.D. in Management Sciences at Jayoti Vidyapeeth Women's University (JVWPU), Rajasthan, with a research focus on data-driven management analytics. Her work encompasses a tangential involvement of Data Science and Mining techniques.

Sheikh Amir Fayaz received an MCA and a Ph.D. from the University of Kashmir. He has over three years of teaching experience and has guided six projects at the master's level. Currently, he is working as an Assistant Professor (C) in the Directorate of IT & SS at the University of Kashmir, J&K. His areas of interest include Machine Learning, Data Mining, and AI.

Dr. Majid Zaman earned his Ph.D. from the University of Kashmir in 2009 and currently serves as a Scientist E in the Department of Information Technology at the University of Kashmir, J&K, India. With over 20 years of teaching experience, he has guided more than 30 master's projects, along with 8 Ph.D. and 3 M.Phil. scholars. His research interests include Data Analytics, Data Mining, Data Science, and Machine Learning.

Dr. Waseem Jeelani Bakshi is an Assistant Professor in the Department of Computer Science and Engineering at the University of Kashmir. He holds a Ph.D. in Computer Sciences and has a strong research focus on Data Science and Machine Learning. His work includes contributions to Data Integration, Predictive Analytics, Big Data Processing and he has published in reputed journals and conferences. Dr. Bakshi is actively involved in academic and research initiatives, fostering innovation and industry collaboration in the field of computer science.

How to cite this paper: Nishit Kaul, Sameer Kaul, Bharti Bhat, Sheikh Amir Fayaz, Majid Zaman, Waseem Jeelani Bakhsi, "Geno-Dwarf-ML: Structural Analysis of Machine Learning Techniques for Genetic Dwarfism Detection", International Journal of Intelligent Systems and Applications(IJISA), Vol.17, No.6, pp.33-43, 2025. DOI:10.5815/ijisa.2025.06.03