UNSUPERVISED MACHINE LEARNING INSIGHTS INTO HEREDITARY DISORDERS THROUGH GENOMIC PROFILING

[1]MARYAM NAGEEN, [2]SYED SAQIB RAZA RIZVI, [3]MUHAMMAD FURQAN

Department of Artificial Intelligence, Department of Computer Science, Department of Software Engineering, Capital University of Science and Technology, Islamabad, Pakistan
E-mail: [1]mariyam3007@yahoo.com, [2]syed.saqibraza@cust.edu.pk, [3]Muhammad.furqan@cust.edu.pk

Abstract- Through genomic profiling researchers identify unique genetic sequences which exist specifically in each family structure. Genomic clustering methods in this proposed analysis distinguish separate genetic groupings or patterns which demonstrate strong inheritance connections between parents and children's genomic data. The clustering model demonstrates strong evidence for its reliability and precision by showing both familiar genetic inheritance patterns alongside unexpected genotypic outcomes such as (TT) whereas certain children illustrate possible de novo or non-Mendelian genetic inheritance patterns. These genomic anomalies compound the inherited genetic tendencies and provide unique perspectives into uncommon genomic patterns enabling exceptional health predictions of cystic fibrosis and achondroplasia. Machine learning algorithms reveals methods to discover patterns inside complex genomic datasets and generate meaningful predictions. These findings support individualized treatment through cantered clinical solutions based on genomic features while tracking family genetic profiles. The approach enables early disease detection followed by targeted treatments while developing preventive strategies for reducing risk. Furthermore, those findings can aid the genetic counseling and foster the development of own family-unique healthcare strategies, establishing a new healthcare paradigm that adapts to the genetic characteristics of each member of the family. The integration of this genomic clustering version into medical practices can revolutionize healthcare via bridging the gap among genomics and real-world medicine, contributing to advancements in predictive healthcare, population genetics, and precision medicine research.

Keywords - Genomic Profiling, Unsupervised Learning (USL), Genomic Clustering, Personalized Healthcare, Autosomal Dominant, Autosomal Recessive

I. INTRODUCTION

A one-size-fits-all approach from the traditional clinical environment has become less efficient since healthcare is evolving toward more personalized care models. This study reveals how genetic insights pair with Machine Learning (ML) to transform personalized healthcare through precision medication alongside genetic-based therapy approaches. [1]. The union of machine learning algorithms with genomic statistics enables pioneers to unravel complex genetic-to-health outcomes connections. This research pursues three main objectives including the development of improved diagnostic and prognostic capabilities and tailored therapies designed for individual patient genetic profiles. [2]. The practice of Machine Learning in genomic precision medicine shows promise to advance customized healthcare through comprehensive genetic disorder understanding [3]. The research focuses on the potential to use Machine Learning algorithms that analyze genetic information for reshaping clinical operations toward individualized and efficient patient treatment systems. Through profound scientific technological advances in genetic precision medication ML achieves new capabilities together with scientific knowledge. [4].

The Human Genome Project (HGP) reached its termination in 2003 and delivered humanity's complete first genomic map.This milestone laid the groundwork for future studies into the links among genetic variations and health, supplying insights into the complexity of human genetics. The HGP stands as one of the most remarkable exceptional scientific achievements. This global collaboration aimed to analyze the entirety of human DNA, or genome, in depth. Its defining achievement—creating the first sequence of the human genome—offered crucial knowledge about the human blueprint, which has enhanced the study in human biology and progressed scientific practices [5]. Figure 1 presents Image of Single Nucleotide Polymorphism in a Human Genome using transmission electron microscope.

Human DNA is built like a microscopic twisted ladder known as a double helix. This iconic shape is formed by two long strands of sugar and phosphate molecules linked together, creating the ladder's backbone. Each rung of the ladder is formed by a pair of nitrogenous bases, which derive in four varieties: adenine (A), guanine (G), cytosine (C), and thymine (T). These hold the key to genetic information in the form of human DNA. Figure 2 presents microscopic double helix from human DNA.

Later, an age of data generation was ushered in by the introduction of high-throughput sequencing technology, which facilitated the creation of massive genetic data databases [6]. When these datasets developed, it became evident how limited existing analytical approaches were, necessitating a paradigm shift. ML has an enormous ability to recognize complex pat- terns and extract important information from these massive genomic data. The merger of ML algorithms with genomics recognized each person's

unique genetic composition and offered a new frontier in customized medicine for individualized diagnostics, medications, and treatments [7].

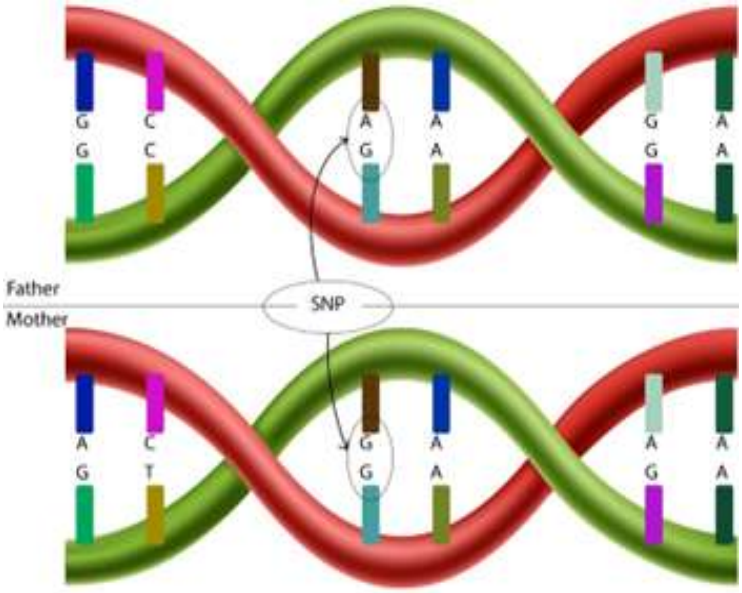


Figure 1: Image of Single Nucleotide Polymorphism in a Human Genome

The key application of ML techniques for genetic research including the development of prediction models for hereditary diseases, and genomic profiling, introduced significant advances in the field [8]. This evolutionary process is far from over since ongoing efforts are made to improve interpretability, optimize algorithms, and handle ethically and privacy related issues [9]. To harness the capacity of ML and comprehend the vision of genomic precision medication, the idea of "genomic profiling" plays a vital role [10]. This involves examining an individual's entire genetic composition, inclusive of DNA sequences, gene expressions, and genetic changes.

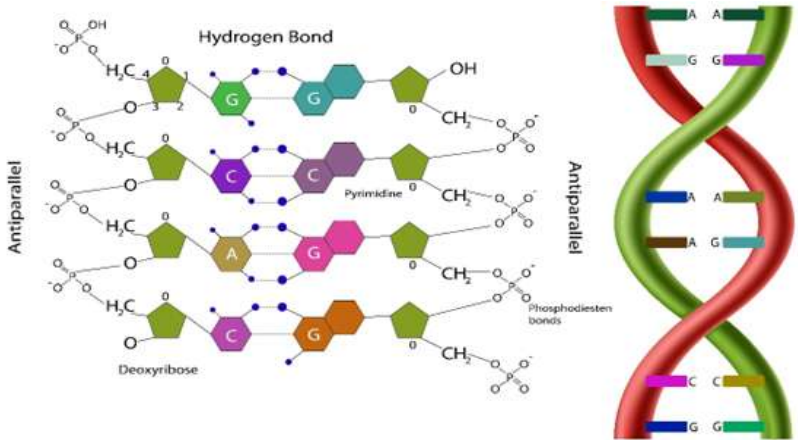


Figure 2: Double Helix Human DNA Structure

Genomic profiling serves because the cornerstone of the personalized healthcare, uncovering the unique genetic pattern that are linked to fitness and disease [3]. Moreover, machine learning algorithms offer a complete framework for extracting significant insights and identifying the patterns inside the huge amount of complex data of genetic information. Expanding beyond the intersection of genomics and ML techniques, the concept of "predictive modeling" emerges as a key approach for developing the models that predict health outcomes based on an individual's genetic information [8].

In the realm of unsupervised mastering (USL), clustering algorithms are involved in uncovering the natural pattern in genomic data [9]. The discovery of natural patterns in genomic data relies on clustering algorithms within unsupervised mastering frameworks [9]. Grouping procedures based on genomic properties allow identification of genetic behavior clusters that match these traits between participants. Inside familial genetics these clustering methods help

identify subpopulations that display unique genetic characteristics and patterns of inherited features. This methodology shows diffused similarities and differences between family genetic data to provide new understanding about complex genomic networks within families. Genomic pattern discoveries push our ability to understand inherited traits together with gene-family interactions using genetic predispositions. Research toward identifying genetic markers of inherited medical issues will drive effective targeted genomic medical practices in the future. [10].

II. MATERIALS & METHODS

There are two main phases in the proposed model: first it processes a dataset followed by data preprocessing. Exploratory data analysis follows preprocessing to deliver valuable information about the records. Feature selection methods identify crucial clustering factors following initial feature selection. A subsequent unsupervised learning algorithm executes the clustering process of selected information. To measure cluster quality the results, undergo evaluation. Document analysis to detect genetic patterns leads to subgroup identification of individuals through genetic profile comparisons. The proposed model is depicted in Figure 3.

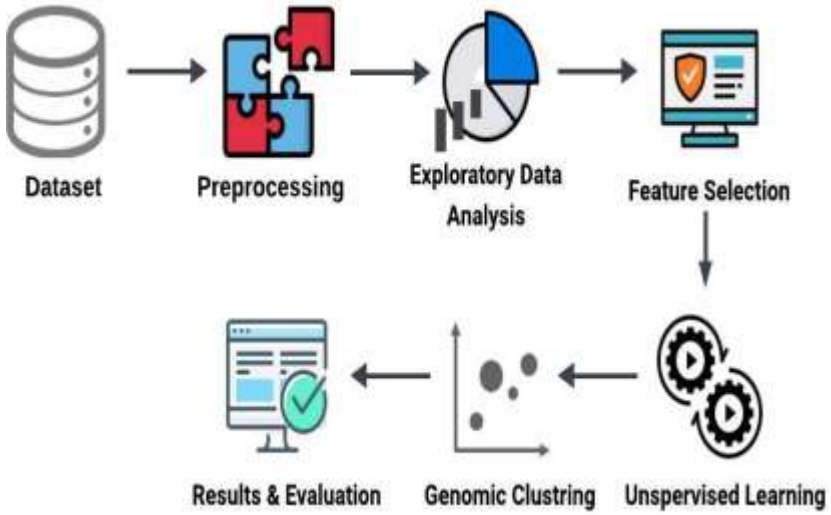


Figure 3: Proposed USL-based Model for Genomic clustering

A. Dataset

Viewed through the scope of whole-genome sequencing, the "Family of Five - Genome Dataset" on Kaggle offers a notable and designated approach into the complicated network of familial genetic relationships, presenting beneficial insights into the hereditary connections and genetic interaction within a circle of family unit. This dataset comprises 5 whole genomes—two from the father and mother and 3 from their child's—observed by means of detailed annotations. The widespread complexity of human genetic information evident in the fact that every genome, representing a single instance, comprises over three billion DNA base pairs.

B. Preprocessing

Three essential characteristics are present in each of the five distinct CSV files representing the five family members that make up the "Family of Five - Genome Dataset" on Kaggle. "Position" pinpoints the location on the genetic map, "Chromosomes" are categorized by 25 unique labels, and the crucial "Genotype" reveals the genetic makeup at each position. However, the "ID" attribute, acting solely as a unique identifier, was deemed superfluous for the intended analysis and discarded. For example, in the "Father Genome," the "Chromosomes" attribute had 25 distinct both numerical and nominal values. The nominal representations were transformed into numerical values. This data harmonization process is applied to standardize this attribute. The attribute "Genotype" comprises 20 distinct values, ranging from the most frequent "GG" with 111990 occurrences to the least frequent "DI" with 33 occurrences. Duplicate data removal was implemented via a two-step process. First, unique identifiers within each file were compared, and redundant instances were eliminated. Subsequent analysis employed checksums to identify identical genetic sequences across files, further streamlining the data for efficient downstream analyses. After the removal of duplicates, the 601802 instances were reduced to 598371 instances in the "Father Genome". The "Genotype" value "X" is assigned to 6616 unlabelled instances. Similarly, the preprocessing was applied to all files of the other four members i.e. "Mother Genome", "Child 1 Genome", "Child 2 Genome" and "Child 3 Genome". The missing values are replaced by mean and mode. This refinement and preprocessing of the dataset were necessary for robust exploration via USL techniques, paving the way for illuminating the intricate genetic relationships woven within the family's genomes.

Figure 4 depict the Attributes of Mother and Father Genome.

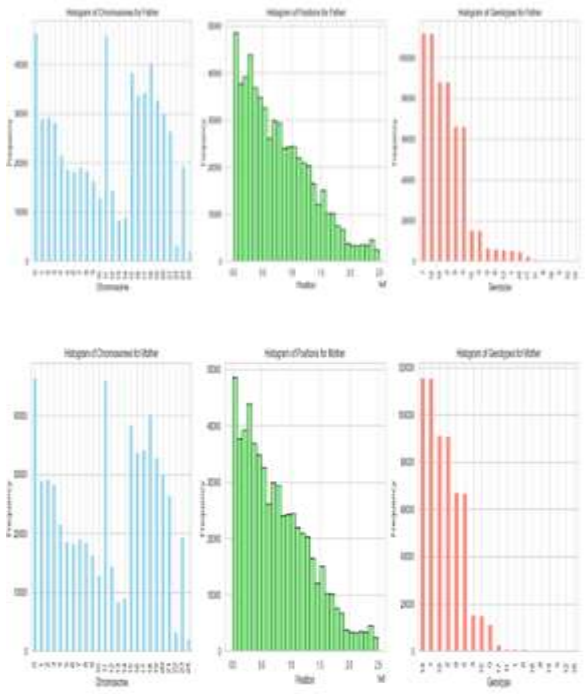


Figure 4: Attributes of Mother and Father Genome

C. Exploratory Data Analysis

After the preprocessing, all the Family of Five genomes were preliminarily analyzed before the subsequent phases. The "Father Genome" is mentioned as an example. The attribute "Chromosomes" has a minimum value of 1 and a maximum value of 25 with a mean of 9.567 and a standard deviation of 6.466. The attribute "Chromosomes" with frequency is illustrated in Figure 5.

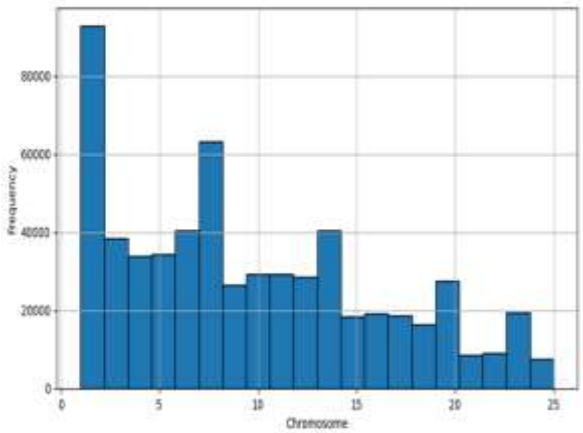
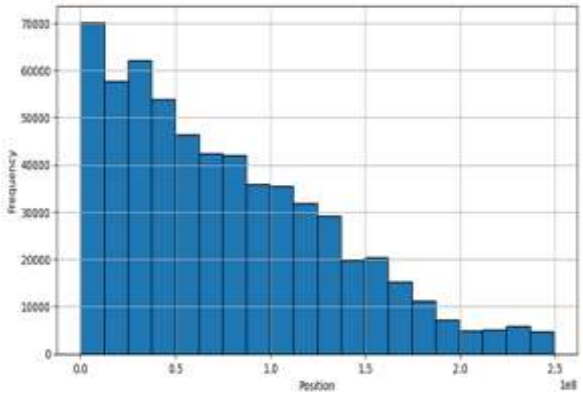


Figure 5: Histogram Attribute "Chromosomes"

The "Position" attribute has to have a minimum value of 3 and a maximum value of 299218992 with a mean equal to 76382309.99 and a standard deviation of 56988227.065. The attribute "Position" with frequency is illustrated in Figure 6.



The attribute "Genotype" is nominal as discussed earlier. The unique value count of the attribute "Genotype" is depicted in Figure 7.

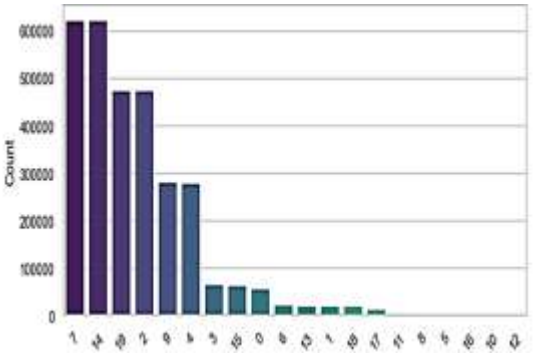


Figure 7: Count of Unique Values in Attribute "Genotype"

D. Feature Selection

Correlations measure the strength and direction of linear relationships between two variables. Knowing how variables are correlated can help to understand how they influence each other. High correlations can suggest potential predictive power. The correlations in the dataset in the form of a heatmap are depicted in Figure 8.

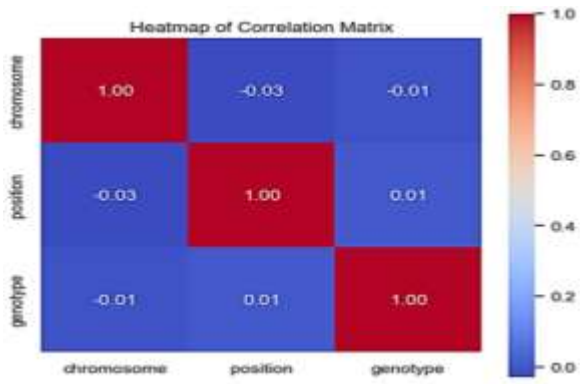


Figure 8: Heatmap to Understand Linear Relationship between Variables

E. Unsupervised Learning

In this data-driven field, unsupervised learning techniques are an effective tool for revealing hidden patterns and structures in large, unlabelled datasets [11]. Unsupervised approaches reveal innate

correlations and clusters among genes and genomic regions, in contrast to supervised learning, where the algorithm is guided by labelled instances [12]. The approach works well for tasks such as finding new gene clusters, locating functional modules, and even detecting genomic markers that may be linked to a disease. Unsupervised learning opens the door for more focused study by revealing hitherto unidentified connections, which might result in advances in condition diagnosis and personalized medication [13-15] Because it can glean insights from the genome's unadulterated symphony, unsupervised learning is a vital tool for deciphering the mysteries of life encoded in its complex genetic code.

F. Genomic Clustering

Genomic clustering is performed on the genomic data of all the family members. The experimental settings for all the family members' genome data were kept identical. The dataset is split into 80 % of training and 20 % test set. The results were obtained using Simple K-means where k = 6 and Euclidian distance. The objective function J is used to calculate the total within-cluster variance. It measures how tightly the data points are grouped in a cluster. This algorithm will try to minimize the J, it finds clusters where data points are close to their own centroids.

$$J = \sum_{i=1}^{k} \sum_{j=1}^{ni} \| xi - \mu i \|^2$$

Where 'k' signifies the number of clusters and 'ni' are the number of those specific data points in the cluster ' I'. The centroid of that cluster which calculates the mean of all data points, represented by μi. This initialize the clusters centroids randomly{ μ1, μ2, …….. μn}, and each point is assigned to the closet centroid cluster Ci. The mean of all points which is assigned to every cluster is used to update the centroid.

$$Ci = \{ x: \| x - \mu i \|^2 \leq \| x - \mu j \|^2, \forall j \neq i \}$$

$$\mu i = \frac{1}{ni} \sum_{x \in Ci} x$$

This whole process of updating of centroid is repeated until no longer is found or change found is below the threshold.

III. RESULTS AND EVALUATION

The "Father Genome" is illustrated in this section as an example. The total number of instances in the "Father Genome" training set is 478696 instances. During the training, the sum of square error within the cluster was recorded as 8.7*104. The training results are illustrated in Table 1.

The clustering analysis discovered 5 distinct clusters, every cluster varying numbers of instances, as specified in Table 1. First Cluster contains 724,255 instances, second Cluster contains 537,500 instances, third Cluster contains 328,386 times, forth Cluster contains 617,403 times, and fifth Cluster contains 129,709 instances. The range of instances of data in every cluster, reflecting the underlying complexity of the genetic facts, with a few clusters containing larger datasets whilst other are smaller. Notably, the model performed 100% accuracy in training, indicating excellent performance in distinguishing between the clusters. The accurate classification throughout all clusters demonstrates the robustness and reliability of the model in identifying genetic patterns. These findings emphasize the effectiveness of the model and its ability for identifying complex genetic relationships, providing a promising basis for destiny applications in genomic studies and personalised medicine. Clustered Instances Test Results shown in Figure 9.

Cluster #	Instances	Chromosomes	Position
Cluster 0	724255	10.767260	4.360147e+07
Cluster 1	537500	12.223165	1.179942e+08
Cluster 2	328386	13.952984	1.623886e+08
Cluster 3	617403	9.816645	7.928058e+07
Cluster 4	129709	5.667901	2.190636e+08
Cluster 5	717278	11.894697	1.357274e+07

Table 1: Father Genome Training Results

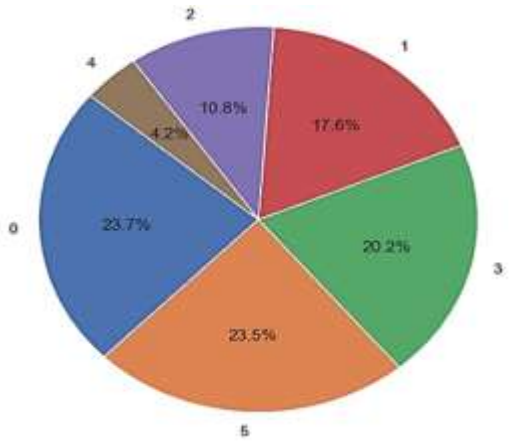


Figure 9: Clustered Instances Test Results

IV. DISCUSSION AND FINDINGS

The results demonstrated that the suggested model we used had a 100% accuracy rate on the test set is far higher than the baseline methods that were previously

offered. Thecluster results of all family members' data is depicted in Table 2.

This result shows that in the first cluster father pass "G" allele and mother can pass on either the "A" or "G" allele to her children. While child-1 inherits ‘A’ allele from the mother and ‘G’ from the father, child-1 is heterozygous (AG) and chil-2 and child-3 both inherited from mother and father so they are homozygous dominant (GG), which means they will have the dominant trait.

Cluster #	Father	Mother	Child -1	Child -2	Child -3
Cluster 0	GG	AG	AG	GG	GG
Cluster 1	CC	CT	CC	TT	TT
Cluster 2	AG	GG	AG	GG	GG
Cluster 3	AA	AA	AA	TT	TT
Cluster 4	GG	AG	GG	AA	AA
Cluster 5	AA	AA	AA	TT	TT

Table 2: Genome Training Results of Clusters

This result shows that in the first cluster father pass "G" allele and mother can pass on either the "A" or "G" allele to her children. While child-1 inherits ‘A’ allele from the mother and ‘G’ from the father, child-1 is heterozygous (AG) and chil-2 and child-3 both inherited from mother and father so they are homozygous dominant (GG), which means they will have the dominant trait. If the gene is autosomal dominant, then only one copy of the dominant allele ‘G’ is required to express the trait. Child-2 and Child-3 (GG) would both show the dominant trait but Child-1 (AG) express the trait of one dominant allele ‘G’. In this cluster Father (GG) and Mother (AG) are both carriers of the gene.

If the disease is autosomal dominant, all three children (Child-1, Child-2, and Child-3) will show the disease or trait because only one copy of the altered gene is enough to cause the disorder and an affected individual has a 50% chance of passing the altered gene to each child, regardless of the child's gender. Autosomal dominant disorders include Neurofibromatosis Type-1 (NF1), which is a common genetic disorder originate tumors to form on the nerve tissue, Huntington's disease is a neurodegenerative disorder causes progressive motor dysfunction, cognitive decline, and psychiatric issues, Achondroplasia is a form of dwarfism, resulting from a mutation in FGFR3 gene, and Marfan syndrome is a connective tissue disorder that affect the heart, vessels, eyes, blood, and skeleton [21].

If the gene is autosomal recessive, two copies of the recessive allele ‘A’ are required to explicit the trait. Child-1 (AG) inherits one ‘A’ from the mother and one ‘G’ from the father, so they may be heterozygous and a carrier and no longer showing the ailment however child-1 can pass the allele. Child-2 and child-3 (GG) inherit the ‘G’ allele from both parents and do not specific the disease but they can also be the carries however do not express the trait. This pattern regularly skips generations, with carriers passing the gene without expressing the disease themselves. Examples of autosomal recessive problems encompass cystic fibrosis which is a genetic problem affecting the lungs, pancreas, liver, and intestines, causes due to the mutations inside the CFTR gene, Tay-Sachs Disease is a neurodegenerative disorder, Phenylketonuria (PKU), which is a metabolic sickness in which the human body cannot break down the amino acid phenylalanine, leading to cognitive disabilities if untreated, and sickle cellular Anemia which is blood disorder wherein red blood cells tackle a sickle form, leading to blockages in blood vessels[18].

In the second cluster Father (CC) is a Homozygous and he does now not have the recessive allele and Mother (CT) is Heterozygous and contains one normal allele ‘C’ and one recessive allele ‘T’. Child-1 (CC), Inherited the ‘C’ allele from both mother and father so Child-1 will be unaffected. Child-2 (TT) and Child-three (TT) Inherited most effective the ‘T’ allele from the mother and no allele from the father which indicates a de novo mutation inside the stated genotypes. Diseases following this inheritance sample are categorized as autosomal recessive disorders. For a child to be affected, that one need to inherit two copies of the recessive allele (TT). Expected ailments consists of Cystic Fibrosis, Sickle Cell Anemia, and Tay-Sachs Disease. This study [17] discusses the genotypes inside the GJB2 gene, where CC represents the homozygous genotype, CT represents the heterozygous carrier, and TT the homozygous mutant related to the hearing loss.

Third cluster is consistent with autosomal dominant inheritance, ‘A’ allele is associated with the dominant trait or disorder. In this cluster Father (AG) is a Heterozygous who is the carrier of one normal allele ‘G’ and one dominant allele ‘A’ and Mother (GG) is Homozygous for the normal allele without dominant trait. Child-1 (AG) Inherited the ‘A’ allele from the father and the G allele from the mother which shows Child-1 affected by the dominant trait, Child-2 (GG) and Child-3 (GG) Inherited the G allele from mother and father so they're not effected. In such conditions, best one copy of the altered gene (A) is enough to cause the disease. These diseases include Achondroplasia, it is a form of short-limbed dwarfism resulting from mutations within the FGFR3 gene. Individuals with one mutated replica (AG) showcase

the circumstance, at the same time as those with regular copies (GG) [22]. Marfan Syndrome is a disorder often resultant from mutations in the FBN1 gene, Huntington's Disease is a neurodegenerative disorder caused by expansions in the HTT gene. A study [16] discusses how specific mutations in the FGFR3 gene lead to achondroplasia, an autosomal dominant disorder.

This genotypic distribution aligns with an autosomal dominant inheritance sample, where the presence of a dominant allele 'A' results in the expression of the related trait or disease. This pattern is found in various genetic conditions, including achondroplasia, that's connected to the mutations within the FGFR3 gene.

In forth cluster a few exciting questions arises Under ordinary Mendelian inheritance policies:

- If each of the parent are (AA), the child must inherit one A allele from each, making all offspring AA.
- The presence of (TT) genotypes in the Child-2 and Child-3 suggests both a de novo mutation, a potential genotyping error, or some form of non-Mendelian inheritance.

If the (TT) genotype arose due to spontaneous mutations in each alleles at some point of early development in Child-2 and Child-3, this can cause genetic problems as a result of homozygous mutations in those specific genes. Expected Diseases with autosomal recessive inheritance appear if the mutation is on a gene critical for some certain traits (e.g., metabolic problems or rare syndromes). Non-Mendelian Inheritance consist of two mechanisms, first one is Uniparental Disomy (UPD) occurs wherein a child inherits two copies of a chromosome or a part of a chromosome from one parent. This is also linked to disorders like Prader-Willi syndrome or Angelman syndrome. Second mechanisms is Somatic Mutations or Mosaicism might also have took place post-zygotically (at some stage in development), leads the differences between germline and somatic cells.

The diseases that would align with this state would probably be uncommon rare genetic disorders. If the mutations affect specific genes, some illnesses might consist of the following disorders like achondroplasia or hypochondroplasia, usually autosomal dominant, however if mutations occur de novo instances are possible. If mutations spontaneously take place in the CFTR gene, it could lead to cystic fibrosis. Smith-Magenis Syndrome (RAI1 Gene) related to de novo mutations. Rett Syndrome (MECP2 Gene) takes place due to spontaneous mutations within the MECP2 gene.

The presence of TT genotypes in Child-2 and Child-3, in which the parents having (AA) genotypes, is exceptionally unusual beneath Mendelian rules. The most probably [19] diseases are linked to the such mutations fall into the class of uncommon rare genetic disorders, frequently affecting genes like FGFR3 will cause excessive cell proliferation, leading to most cancer and skin overgrowth [paper: The paradox of FGFR3 signaling in skeletal dysplasia: Why chondrocytes growth arrest while other cells over proliferate Author: Pavel Krejci], CFTR or others associated with metabolic or developmental syndromes. Further genetic checking out or validation of the records would be needed for the confirmation in this case.

Autosomal recessive inheritance is also being followed by the fifth cluster, as well as carries the instances of non-Mendelian or mutations. Father (CG) is a Homozygous for the 'G' allele, so he can only pass on G allele and Mother (AG) is a Heterozygous who carries one 'A' allele and one 'G' allele. Child-1 (GG) can Inherit 'G' from both parents and Child-2(AA) and Child-3(AA) Inherited 'A' from the mother which increases the possibility of De novo mutation in the father's germline or Uniparental disomy. Diseases fall under autosomal recessive disorders getting two copies of the recessive allele (AA) are: Sickle Cell (Behl n.d.)Disease (HBB Gene) caused by mutations in the HBB gene, Cystic Fibrosis (CFTR Gene) caused by mutations in the CFTR gene as it is realized that individuals who were homozygous for the common F508del variant varied considerably in the severity of lung disease [20], Metabolic Disorders like phenylketonuria (PKU) or Tay-Sachs disease, and Rare Neurological or Developmental Syndromes.

If the Single Nucleotide Polymorphisms (SNPs) are located on different chromosomes or some distant genomic regions, they are not linked, occurred in cluster 4 and 6. The probability of them being inherited collectively at some stage in meiosis may be very low due to the excessive possibility of crossover activities occurring among them in the course of chromosomal recombination. Parallel inheritance of unrelated alleles can also contain the Multiple loci inside genome self-sufficiently produce (AA) or (TT) genotypes in a Mendelian way or Genotype patterns is probably preserved because of evolutionary conservation of the inheritance mechanisms.

IV. CONCLUSION

In conclusion,clustering analysis on the "Family of Five" dataset exposed many significant patterns of inheritance together with important genetic discoveries. Genetic alterations linked to dominant inheritance often make themselves apparent through the simultaneous occurrence of offspring displaying dominant condition and carrier phenotypes in other offspring. The occurrence of (TT) genotypes in particular observed cases indicated potential de novo

mutations or genotyping mistakes or uniparental disomy acting as mechanisms commuting Mendelian inheritance patterns. The genetic transmission proved complicated by these anomalies which offered evidence for associations with diseases like cystic fibrosis and achondroplasia. This research detailed the multilayered relationship between heritable patterns combined with newly generated mutations alongside unusual genetic mechanisms that guide the understanding of genetic issues and familial traits. The discovered familial genomic patterns require further research validation to explore de novo mutations while deepening our comprehension of such inheritance dynamics. Precision medicine development reaches new milestones through improvements in methods used to identify and manage uncommon inherited diseases resulting in individualized therapeutic opportunities.

REFERENCES

[1] Ahmed, Z., Zeeshan, S., Mendhe, D., & Dong, X. 2020. "Human gene and disease associations for clinical-genomics and precision medicine research. ,." Clinical and translational medicine 10(1), 297-318.

[2] Anklam, E., Bahl, M. I., Ball, R., Beger, R. D., Cohen, J., Fitzpatrick, S., ... & Slikker Jr, W. 2022. "Emerging technologies and their impact on regulatory science." Experimental Biology and Medicine 247(1).

[3] Behl, T., Kaur, I., Sehgal, A., Singh, S., Albarrati, A., Albratty, M., ... & Bungau, S. n.d. "The road to precision medicine: Eliminating the “one size fits all” approach in Alzheimer’s disease." Biomedicine & Pharmacotherapy, 153, 113337.

[4] Brendel, M., Su, C., Bai, Z., Zhang, H., Elemento, O., & Wang, F. 2022. "Application of deep learning on single-cell RNA sequencing data analysis: a review. Genomics, Proteomics & Bioinformatics, 20(5), 814-." Genomics, Proteomics & Bioinformatics 20(5), 814-835.

[5] Chafai, N., Bonizzi, L., Botti, S., & Badaoui, B. 2023. "Emerging applications of machine learning in genomic medicine and healthcare." Critical Reviews in Clinical Laboratory Sciences 1-24.

[6] Cutting, . Sharma · G.R. 2021. " The genetics and genomics of cystic fibrosis." PubMed Central (PMC) 11.

[7] García-Sancho, M., Leng, R., Viry, G., Wong, M., Vermeulen, N., & Lowe, J. 2022. "The Human Genome Project as a singular episode in the history of genomics." Historical Studies in the Natural Science 52(3),320-360.

[8] Gill, A. Y., Saeed, A., Rasool, S., Husnain, A., & Hussain, H. K. 2023. "Revolutionizing Healthcare: How Machine Learning is Transforming Patient Diagnoses Comprehensive Review of AI's Impact on Medical Diagnosis." Journal of World Science, 2(10), 1638-1652.

[9] J E Spence 1, R G Perciaccante, G M Greig, H F Willard, D H Ledbetter, J F Hejtmancik, M S Pollack W E O'Brien, A L Beaudet. 1988. "Uniparental disomy as a mechanism for human genetic disease." Elsevier 10.

[10] Lauschke, Qingyang Xiao and Volker M. 2021. " The prevalence, genetic complexity and population specific founder effects of human autosomal recessive disorders." NPJ genomic medicine 7.

[11] Legare, Janet M. n.d. Achondroplasia (Synonym: FGFR3-Related Achondroplasia). MD School of Medicine and Public Health University of Wisconsin].

[12] Lewis1, Rueben G., and Brittany Simpson2. n.d. Genetics, Autosomal Dominant. 1 Nova Southeastern University, 2 Cincinnati Children's Hospital.

[13] Li, C., Pan, Y., Zhang, R., Huang, Z., Li, D., Han, Y., ... & Kelly, T. N. 2023. "Genomic innovation in early life cardiovascular disease prevention and treatment." Circulation Research 132(12), 1628-1647.

[14] Lorkowski, J., Kolaszyńska, O., & Pokorski, M. 2021. "Artificial intelligence and precision medicine: A perspective. In Integrative Clinical Research." Cham: Springer International Publishing 1-11.

[15] R Shiang 1, L M Thompson, Y Z Zhu, D M Church, T J Fielder, M Bocian, S T Winokur, J J Wasmuth. 1994. "Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia." CELL 674(94)90302-6.

[16] Rasool, S., Husnain, A., Saeed, A., Gill, A. Y., & Hussain, H. K. 2023. "Harnessing Predictive Power: Exploring the Crucial Role of Machine Learning in Early Disease Detection." JURIHUM: Jurnal Inovasi dan Humaniora 1(2), 302-315.

[17] Renner, H., Schöler, H. R., & Bruder, J. M. 2021. " Combining Automated Organoid Workflows with Artificial Intelligence-Based Analyses: Opportunities to Build a New Generation of Interdisciplinary High-Throughput Screens for Parkinson's Disease and Beyond." Movement Disorders 6(12), 2745-2762.

[18] Sindelar, R. D. 2024. " Genomics, other “OMIC” technologies, precision medicine, and additional biotechnology-related techniques." In Pharmaceutical Biotechnology: Fundamentals and Applications 18(6), 209-254.

[19] Somayeh Ebrahimkhani, Golnaz Asaadi Tehrani. 2021. "Evaluation of the GJB2 and GJB6 Polymorphisms with Autosomal Recessive Non-Syndromic Hearing Loss ." Iranian Journal of Otorhinolaryngology 33(115):79–86.

[20] Thirunavukarasu, R., Gnanasambandan, R., Gopikrishnan, M., & Palanisamy, V. 2022. " Towards computational solutions for precision medicine based big data healthcare system using deep learning models: A review." . Computers in Biology and Medicine 106020.

[21] Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. 2021. "The ethics of algorithms: key problems and solutions." Ethics, Governance, and Policies in Artificial Intelligence 97-123.

[22] Yan, J., & Wang, X. 2022. "Unsupervised and semi-supervised learning: the next frontier in machine learning for plant systems biology." The Plant Journal 111(6), 1527-1538.

★★★